CUK-H03678-7-P024845

ANNUAL REVIEW OF MEDICINE:

Selected Topics in the Clinical Sciences

⑦

UNIVERSITY COLLEGE OF MEDICINE
LIBRARY
C. U.

1

P, 24845

EDITORIAL COMMITTEE (1977)

C. H. COGGINS
Harvard Medical School

W. P. CREGER
Stanford University

E. W. HANCOCK
Stanford University

K. J. ISSELBACHER
Massachusetts General Hospital

S. L. KATZ
Duke University

E. J. SACHAR
Albert Einstein College of Medicine

G. T. SHIRES
Cornell University

M. SIPERSTEIN
University of California Service
Veterans Administration Hospital,
San Francisco

R. T. SMITH
The University of Florida

Responsible for the organization of Volume 28
(Editorial Committee, 1975)

C. H. COGGINS
Harvard Medical School

W. P. CREGER
Stanford University

E. W. HANCOCK
Stanford University

D. C. SABISTON JR.
Duke University

E. J. SACHAR
Albert Einstein College of Medicine

J. P. SANFORD
The University of Texas

M. SIPERSTEIN
University of California Service
Veterans Administration Hospital,
San Francisco

R. T. SMITH
The University of Florida

L. W. WANNAMAKER
The University of Minnesota

Assistant Editor JULIE FAGAN

ANNUAL REVIEW OF MEDICINE:

Selected Topics in the Clinical Sciences

WILLIAM P. CREGER, *Editor*
Stanford University School of Medicine

C. H. COGGINS, *Associate Editor*
Harvard Medical School

E. W. HANCOCK, *Associate Editor*
Stanford University School of Medicine

VOLUME 28

1977

ANNUAL REVIEWS INC. 4139 EL CAMINO WAY PALO ALTO, CALIFORNIA 94306

ANNUAL REVIEWS INC.
Palo Alto, California, USA

COPYRIGHT © 1977 BY ANNUAL REVIEWS INC., PALO ALTO, CALIFORNIA.
ALL RIGHTS RESERVED. No part of this book may be reproduced in any form or by any means without permission in writing from the publisher.

International Standard Book Number: 0-8243-0528-0
Library of Congress Catalog Card Number: A51-1659

Annual Reviews Inc. and the Editors of its publications assume no responsibility for the statements expressed by the contributors to this Review.

REPRINTS

The conspicuous number aligned in the margin with the title of each article in this volume is a key for use in ordering reprints. Available reprints are priced at the uniform rate of $1 each postpaid. The minimum acceptable reprint order is 10 reprints and/or $10.00, prepaid. A quantity discount is available.

1800

PRINTED AND BOUND IN THE UNITED STATES OF AMERICA

CONTENTS

ALLERGY AND IMMUNOLOGY:

The Association of Genes in the Major Histocompatibility Complex and Disease Susceptibility, *Takehiko Sasazuki, F. Carl Grumet, and Hugh O. McDevitt* 425

The Current Status of BCG Immunization Against Tuberculosis, *Theodore C. Eickhoff* 411

BCG as Adjuvant Immunotherapy for Neoplasia, *Evan M. Hersh, Jordan U. Gutterman, and Giora M. Mavligit* 489

Structure and Function of Alpha-Fetoprotein, *Thomas B. Tomasi, Jr.* 453

IgA-Associated Glomerulonephritis, *John J. McPhaul, Jr.* 37

ANESTHESIA:

Acupuncture: Context and Critique, *Bernard S. Millman* 223

Malignant Hyperpyrexia, *C. Ronald Stephen* 153

Effects of Anesthetic Drugs on Myocardial Performance in Man, *Robert G. Merin* 75

CARDIOVASCULAR:

The Role of Ethanol in Cardiac Disease, *Timothy J. Regan, Philip O. Ettinger, Bunyad Haider, S. Sultan Ahmed, Henry A. Oldewurtel, and Michael M. Lyons* 393

Ionic Basis of Myocardial Contractility, *Glenn A. Langer* 13

Critique of Coronary Artery Bypass Surgery, *Charles B. Mullins and Kirk Lipscomb* 271

Hypertrophic Cardiomyopathy, *Pravin M. Shah* 235

The Detection of Asymptomatic Coronary Artery Disease, *Victor F. Froelicher, Jr.* 1

Prophylaxis of Infective Endocarditis: A Reevaluation, *James N. Sipes, Robert L. Thompson, and Edward W. Hook* 371

Development of Coronary Collateral Circulation, *Andrew S. Wechsler* 341

DERMATOLOGY:

Pathophysiology of Psoriasis, *John J. Voorhees* 467

Current Problems in Mycosis Fungoides and Sézary Syndrome, *R. K. Winkelmann and W. A. Caro* 251

ENDOCRINOLOGY AND METABOLISM:

The Role of Glucagon in the Endogenous Hyperglycemia of Diabetes Mellitus, *Roger H. Unger and Lelio Orci* 119

Impaired Peripheral Conversion of Thyroxine to Triiodothyronine, *Ralph R. Cavalieri and Basil Rapoport* 57

Diagnosis of Hyperparathyroidism, *Henry G. Bone, III, William H. Snyder, III, and Charles Y. C. Pak* 111

Classification of the Hyperlipidemias, *Richard J. Havel* 195

Paget's Disease of Bone, *Will G. Ryan* 143

GASTROINTESTINAL TRACT:

Viral Agents and Diarrheal Illness, *Herbert L. DuPont, Benjamin L. Portnoy, and Richard H. Conklin* 167

HEMATOLOGY:

Treatment of Idiopathic Thrombocytopenic Purpura (ITP), *Yeon S. Ahn and William J. Harrington* 299

Paroxysmal Nocturnal Hemoglobinuria (PNH) as a Clonal Disorder, *Robert C. Hartmann and Anita B. Arnold* 187

Cold Agglutinins and Cold-Agglutinin Disease, *Michael M. Frank, John P. Atkinson, and James Gadek* 291

Factors in the Pathomechanism of Chronic Lymphocytic Leukemia, *Harald Theml, Richard Love, and Herbert Begemann* 131

INFECTION:

Pertussis in the Adult, *Calvin C. Linnemann, Jr. and Jerome Nasenbeny* 179

Cytomegalovirus (CMV) in the Compromised Host(s), *Robert F. Betts and James B. Hanshaw* 103

The Emergence of Group B Streptococci in Infections of the Newborn Infant, *Bascom F. Anthony and Donald M. Okada* 355

KIDNEYS AND URINARY SYSTEM:

Urate Excretion, *George M. Fanelli, Jr.* 349

Renal Function During and Following Obstruction, *Douglas R. Wilson* 329

Glomerular Filtration of Fluid and Macromolecules: The Renal Response to Injury, *David A. Maddox and Barry M. Brenner* 91

MISCELLANEOUS:

Confidentiality in Medical Practice, *Maurice Grossman* 43

NEOPLASIA:

The Role of Oncogenic Viruses in Neoplasia, *Paul A. Klein and Richard T. Smith* 311

Surgical Adjuvant Chemotherapy, *Marjorie Perloff and James F. Holland* 475

NUTRITION:

Applications and Hazards of Total Intravenous Hyperalimentation, *Stanley J. Dudrick and James M. Long, III* 517

OPHTHALMOLOGY:

Corneal Transplantation, *S. Lance Forstot and Herbert E. Kaufman* 21

PSYCHIATRY:

New Concepts in the Diagnosis of Psychiatric Disorders, *Eli Robins* 67

RESPIRATORY SYSTEM:

Diagnosis and Treatment of Fat Embolism, *Joseph A. Moylan and Merle A. Evenson* 85

Pulmonary Complications of General Surgery, *Alan K. Pierce and Joy Robertson* 211

Diagnosis of Pulmonary Embolism, *G. V. R. K. Sharma, Donald E. Tow, Alfred F. Parisi, and Arthur A. Sasahara* 159

CUMULATIVE INDEX OF CONTRIBUTING AUTHORS, VOLUMES 24–28 529

CUMULATIVE INDEX OF CHAPTER TITLES, VOLUMES 24–28 532

ANNUAL REVIEWS INC. is a nonprofit corporation established to promote the advancement of the sciences. Beginning in 1932 with the *Annual Review of Biochemistry,* the Company has pursued as its principal function the publication of high quality, reasonably priced Annual Review volumes. The volumes are organized by Editors and Editorial Committees who invite qualified authors to contribute critical articles reviewing significant developments within each major discipline.

Annual Reviews Inc. is administered by a Board of Directors whose members serve without compensation.

BOARD OF DIRECTORS
1977

Dr. J. Murray Luck
Founder Emeritus, Annual Reviews Inc.
Department of Chemistry
Stanford University

Dr. Joshua Lederberg
President, Annual Reviews Inc.
Department of Genetics
Stanford University Medical School

Dr. James E. Howell
Vice President, Annual Reviews Inc.
Graduate School of Business
Stanford University

Dr. William O. Baker
President
Bell Telephone Laboratories

Dr. Sidney D. Drell
Deputy Director
Stanford Linear Accelerator Center

Dr. Eugene Garfield
President
Institute for Scientific Information

Dr. William D. McElroy
Chancellor
University of California, San Diego

Dr. John Pappenheimer
Department of Physiology
Harvard Medical School

Dr. Colin S. Pittendrigh
Director
Hopkins Marine Station

Dr. Esmond E. Snell
Department of Microbiology
University of Texas at Austin

Dr. Alvin M. Weinberg
Director, Institute for Energy Analysis
Oak Ridge Associated Universities

Dr. Harriet Zuckerman
Department of Sociology
Columbia University

Annual Reviews are published in the following sciences: Anthropology, Astronomy and Astrophysics, Biochemistry, Biophysics and Bioengineering, Earth and Planetary Sciences, Ecology and Systematics, Energy, Entomology, Fluid Mechanics, Genetics, Materials Science, Medicine, Microbiology, Nuclear Science, Pharmacology and Toxicology, Physical Chemistry, Physiology, Phytopathology, Plant Physiology, Psychology, and Sociology. The *Annual Review of Neuroscience* will begin publication in 1978. In addition, two special volumes have been published by Annual Reviews Inc.: *History of Entomology* (1973) and *The Excitement and Fascination of Science* (1965).

Ann. Rev. Med. 1977. 28:1–12
Copyright © 1977 by Annual Reviews Inc. All rights reserved

THE DETECTION OF ASYMPTOMATIC CORONARY ARTERY DISEASE

❖7217

Victor F. Froelicher, Jr., M. D.
Department of Cardiology, Wilford Hall USAF Medical Center, Lackland AFB, Texas 78239

Modification of the modern western life-style could reduce the epidemic of coronary atherosclerotic heart disease (CAD) (1–5). Since several susceptible generations will be exposed before primary prevention is a reality, it is advisable to evaluate screening methods for detecting asymptomatic CAD. The relative value of techniques for identifying individuals who have asymptomatic CAD should be assessed in order to optimally direct preventive efforts to them. Also, there is a necessity to evaluate individuals in whom sudden incapacitation could compromise public safety. Another concern is to determine the cost effectiveness of the available screening techniques. This chapter reviews some principles of applying screening tests to populations of asymptomatic men.

In order to evaluate a test for a disease, one must demonstrate how well the test distinguishes between individuals with and individuals without the disease (6). Data available from the use of maximal or near maximal exercise testing (7–10) is used to demonstrate how a screening test can be evaluated. Table 1 lists the possible causes of an abnormal electrocardiographic response to an exercise test. CAD is listed first, since it is today's major health problem and obviously is the main diagnostic objective of exercise testing. Other conditions, such as valvular and congenital heart disease, can be detected in better ways and poor testing techniques must be avoided. The other causes listed in Table 1 are hypothetical, including subclinical cardiomyopathies and subclinical myocarditis. These other possible causes do not pose as significant a threat to public health as does coronary atherosclerotic heart disease.

DIAGNOSTIC ENDPOINTS FOR CAD

In order to evaluate exercise testing as a screening test, the populations tested must be divided into CAD and non-CAD groups by independent techniques. Exercise testing, coronary angiography, and clinical events are three methods of separating

Table 1 Possible causes of an abnormal ECG response to exercise testing

1. Coronary atherosclerotic heart disease (myocardial ischemia)
2. Conditions and techniques causing abnormal responses not due to coronary artery disease (false positives)
3. Others (hypothetical)
    - Normal variant
    - Cardiomyopathies
    - Coronary artery spasm
    - Cellular — membrane defect
    - Viral myocarditis
    - Neurogenic imbalance
    - Small-vessel disease
    - Oxygen-hemoglobin dysfunction

populations into disease and non-disease groups. Some of the aspects and limitations of the abnormal electrocardiographic (ECG) response to exercise testing follow below:

1. Nonspecific nature of repolarization abnormalities (ST-TW)—not only due to ischemia, i.e. orthostatic (11), hyperventilation (12), eating (13), drugs (14), and abnormal depolarization (15).
2. Repolarization abnormalities secondary to relative ischemia—hypertrophy (16), valvular and congenital heart disease (17), excessive double product (18), and sudden high workload (19).
3. Technical problems—inadequate recording equipment (20), improper lead system (21), and imprecise interpretation (22).

ST-segment depression is nonspecific; that is, it is not only caused by ischemia. ST-segment depression can be secondary to conditions that cause relative ischemia. As in any type of testing, technical problems can affect the interpretation. These things must be considered when using exercise testing as a diagnostic endpoint. They also partially explain why an abnormal test cannot be equated with CAD.

The most accurate ways of separating CAD and non-CAD groups are coronary angiography and clinical CAD events. Coronary angiography has certain limitations.

It has been demonstrated in numerous studies comparing angiographic and pathologic findings that coronary angiography usually underestimates the pathologic severity of CAD (23). Coronary angiography can be interpreted as normal when severe CAD is present. This can be due to total cutoff of an artery at its origin (24), by diffuse atherosclerotic narrowing of an artery (25), and by proximal left coronary-artery lesions, which can be seen only by special axial views (26). Another limitation of coronary angiography is that spasm of a coronary artery may be transient and missed during the catheterization procedure.

There are some important aspects and limitations of the clinical events and pathological endpoints that are used to separate CAD and disease-free groups. CAD events and clinical CAD can be due to relatively minor lesions. Hemorrhage into nonobstructive plaques can result in total arterial occlusion and death (27, 28). Also, spasm (causing symptoms and ECG changes) has been demonstrated to occur very commonly near relatively minor lesions (29). Pathological studies have shown that approximately 7% of people dying of clinical myocardial infarction have insignificant or no CAD (30). Coronary angiographic studies have shown that some patients with classical angina pectoris (31) and myocardial infarctions (32) can have normal coronary angiograms.

In spite of these limitations, coronary angiography and clinical findings or events are presently the most practical endpoints to separate a population into those with CAD and those without CAD.

SENSITIVITY AND SPECIFICITY

Sensitivity and specificity are the terms used to define how reliably a test distinguishes diseased from non-diseased individuals (6). Sensitivity is the percentage of times a test gives a positive result when the subgroup with the disease is tested. Specificity is the percentage of times the test gives a negative result when the subgroup without disease is tested. This is quite different from the colloquial use of the word "specific." The method of calculating these terms is shown in Figure 1.

The sensitivity of near-maximal or maximal exercise testing can be estimated by analyzing certain groups of patients in studies that have used exercise testing to evaluate hospital or clinic patients. From such studies patients can be selected who have both clinical findings of CAD and abnormal coronary angiography. These patients have been diagnosed as having CAD by two independent techniques: clinical judgment and coronary angiography. Certainly it must be assumed that reasonable methods and criteria have been used by the different investigators. In Table 2 are seven studies listed by the year reported and by the first author and place of origin. The denominators are the number of patients with both clinical CAD and abnormal coronary angiography. The numerators are the number of patients who had clinical CAD, abnormal coronary angiography, and abnormal exercise tests. The sensitivities ranged from 33% to 82%. Bartel et al (38) studied the largest number of patients and their sensitivity of 65% was not much different from the mean sensitivity determined from the other six groups combined (64%). Case reports of individuals who had CAD events immediately following a normal maximal or near-maximal exercise test (27, 40) emphasize the sensitivity limitations of exercise testing.

As previously defined, the specificity of a test is the percentage of those without the disease who have a negative test. It is difficult to find sufficient numbers of

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \times 100 \qquad \text{Relative Risk} = \frac{\dfrac{\text{TP}}{\text{TP} + \text{FP}}}{\dfrac{\text{FN}}{\text{TN} + \text{FN}}}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{FP} + \text{TN}} \times 100 \qquad \begin{array}{c}\text{Predictive value}\\ \text{of an abnormal test}\end{array} = \frac{\text{TP}}{\text{TP} + \text{FP}} \times 100$$

Figure 1 Calculation of sensitivity, specificity, relative risk, and predictive value. (TP = true positives or those with abnormal test and disease; FN = false negatives or those with normal test and with disease; TN = true negatives or those with normal test and no disease; and FP = false positives or those with abnormal test and no disease.)

Table 2 Sensitivity of exercise testing (maximal or near maximal) in patients with clinical CAD and abnormal coronary angiograms

Study (Ref.)	Percentage with abnormal test (sensitivity)	
Mason; John Hopkins (33) (1967)	78	$\left(\frac{38}{49}\right)$
Kasselbaum; Univ. Oregon (34) (1968)	50	$\left(\frac{18}{36}\right)$
Roitman; Univ. Alabama (35) (1970)	80	$\left(\frac{24}{30}\right)$
Ascoop; The Netherlands (36) (1971)	59	$\left(\frac{26}{44}\right)$
McHenry; Univ. Indiana (37) (1972)	82	$\left(\frac{70}{85}\right)$
Bartel; Duke (38) (1974)	65	$\left(\frac{216}{332}\right)$
Borer; NIH (39) (1975)	33	$\left(\frac{13}{39}\right)$
Total	66	$\left(\frac{405}{615}\right)$
	(Mean of percentages = 64)	

individuals without clinical CAD and with coronary angiography, because there are few patients studied with coronary angiography who do not have clinical CAD and do not have abnormal exercise tests. Therefore I have elected to utilize data from follow-up studies. This involves using an endpoint determined from an incidence study as opposed to a prevalence study, and may only approximate the specificity obtained from using coronary angiography. Table 3 lists four follow-up studies in which asymptomatic, apparently healthy men were tested using maximal or near-maximal exercise and then followed for two to six years for the manifestations of CAD. The numerators are subjects who did not develop the manifestations of CAD and had a normal test, and the denominators are all subjects who did not develop CAD over the follow-up period. The calculated specificities of 90% to 92% are surprisingly consistent.

RELATIVE RISK AND PREDICTIVE VALUE OF AN ABNORMAL TEST

When using a screening test, it is also important to know what the relative risk of an abnormal test will be and what its predictive value will be. Figure 1 shows how these terms are calculated. The relative risk is the relative chance of having disease if the test is abnormal, as compared to having disease if the test is normal. Table 4 shows the relative risk for the subjects with an abnormal test of developing CAD, as compared to subjects with a normal test. The risks range from 10 times to 14.3

Table 3 Specificity of exercise testing in four studies that screened asymptomatic men with maximal or near-maximal exercise testing and followed them for manifestations of CAD

Study (Ref.)	Percentage of those without CAD with normal test (specificity)	
Bruce; Univ. Washington (41) (1969) 5 yr	91	$\left(\frac{197}{216}\right)$
Aronow; Long Beach VA Hosp. (42) (1975) 5 yr	92	$\left(\frac{84}{91}\right)$
Froelicher; USAFSAM (8) (1974) 6.3 yr	92	$\left(\frac{1232}{1344}\right)$
Cumming; Univ. Manitoba (43) (1975) 3 yr	90	$\left(\frac{438}{484}\right)$
Total	91	$\left(\frac{1951}{2135}\right)$

Table 4 Results of four follow-up studies using maximal or near-maximal exercise testing to screen asymptomatic men for latent CAD

Follow-up studies (Ref.)	Relative risk of developing CAD over follow-up period of 2–6 yr for those with an abnormal exercise test	Predictive value of an abnormal exercise test (%)	False-positive rate (%)
Bruce (41)	13.6	13.6	86.4
Aronow (42)	13.6	46	54
Froelicher (8)	14.3	20	80
Cumming (43)	10	25	75

times for the abnormal responders; thus the exercise test has been demonstrated to find a subgroup of apparently healthy, asymptomatic men at very high risk of developing CAD.

Predictive value of an abnormal test is the percentage of those with an abnormal test who have the disease. The percentage difference, calculated by subtracting the predictive value from 100% will be the false-positive rate. As shown in Table 4, if the predictive value is 13.6–46%, that percentage of the abnormal responders will have the disease and 54–86.4% of the abnormal responders will not have the disease. Neither the predictive value nor the false-positive rate can be estimated directly from a test's demonstrated specificity or sensitivity. Predictive value and the false-positive rate are dependent upon the prevalence of disease in the population tested. This can be easily explained by a series of simple calculations.

Table 5 Performance of a test with a 60% sensitivity and a 90% specificity in a population with a 1% prevalence of disease

Subjects	Number with abnormal test			Number with normal test	
100 diseased	60	(TP)		40	(FN)
			(Sensitivity)		
9900 nondiseased	990	(FP)		8910	(TN)
			(Specificity)		
Total	1050			8950	

$$\text{Predictive value of an abnormal test} = \frac{\text{TP}}{\text{TP} + \text{FP}} = \frac{60}{1050} = 5.7\%$$

False-positive rate = 100 – 5.7 = 94.3%

Relative risk = 12.7

Table 5 demonstrates how a test with a sensitivity of 60% and a specificity of 90% performs in a population with a 1% prevalence of disease. As previously demonstrated, these values approximate the sensitivity and specificity of maximal or near-maximal exercise testing. Since 1% of the 10,000 men have disease, 100 men have disease. In the middle column are the number of men with abnormal tests and in the far right column are the number with normal tests. Since the test is 60% sensitive, 60% of those with disease will have abnormal tests and are true positives. The remaining 40 have normal tests and are false negatives. Since the test is 90% specific, 90% of the 9900 without disease are true negatives, while the remainder are false positives. To calculate the predictive value, the true positives are divided by all of those with an abnormal test. Table 6 shows the performance of a test with the same 60% sensitivity and 90% specificity in a population with a 10% prevalence

Table 6 Performance of a test with a 60% sensitivity and a 90% specificity in a population with a 10% prevalence of disease

Subjects	Number with abnormal test			Number with normal test	
1000 diseased	600	(TP)		400	(FN)
			(Sensitivity)		
9000 nondiseased	900	(FP)		8100	(TN)
			(Specificity)		
Total	1500			8500	

$$\text{Predictive value of an abnormal test} = \frac{\text{TP}}{\text{TP} + \text{FP}} = \frac{600}{1500} = 40\%$$

False-positive rate = 100 – 40 = 60%

Relative risk = 8.5

of disease. After performing the same calculations as in Table 5, it is obvious that using the test in a population with a greater disease prevalence reduces the false-positive rate.

Table 7 shows the predictive value and the false-positive rate for an abnormal test with a 60% sensitivity and a 90% specificity when used in four populations of differing prevalences. The values were determined as calculated in the two previous tables. It is obvious that as the prevalence of disease increases, the predictive value increases and the false-positive rate decreases. Table 8 shows how a test with a 95% specificity and 95% sensitivity would function in populations with a spectrum of disease prevalences. Such a test would be the near-perfect test and presently there is no such screening technique available. However, the same relationship of disease prevalence and predictive value is again found, confirming that this relationship is inherent in all tests. This is consistent with the Bayes theorem (44) and the predictive model (45).

CORONARY ANGIOGRAPHY AND TREADMILL TESTING AT USAFSAM

The results of cardiac catheterization and treadmill testing in aircrewmen has supported this relationship (46). Two subgroups can be contrasted; one with a low and one with a high prevalence of CAD. Asymptomatic, apparently healthy aircrewmen referred to the USAF School of Aerospace Medicine (USAFSAM) consultation service probably have a CAD prevalence less than the general population because of the selective process to enter and remain on flying status. One hundred and eleven such aircrewmen had coronary angiography only because of an abnormal maximal treadmill test and 30.6% of them were found to have at least one lesion, $\geq$ 50% in a major coronary artery. Thus the predictive value of an abnormal treadmill test in them was approximately 30.6%. In contrast, angiography was also performed, no matter what the result of treadmill testing, in 52 men who were referred because of mild probable angina pectoris and/or a probable remote myocardial infarction. Since this group of men were more similar to a hospital or clinic population, they had a higher prevalence of CAD. Approximately 60% of them had

Table 7 The effect of the prevalence of a disease on the predictive value and false-positive rate of an abnormal response to a test that has a 60% sensitivity and 90% specificity

Prevalence (%)	Predictive value (%)	False-positive rate (%)
1.0	5.7	94.3
5.0	24	76
10.0	40	60
50.0	86	14

Table 8 The effect of the prevalence of a disease on the predictive value and false-positive rate of an abnormal response to the near-perfect test (95% sensitivity, 95% specificity)

Prevalence (%)	Predictive value (%)	False-positive rate (%)
0.1	2	98
1.0	16	84
2.0	28	72
5.0	50	50
10.0	68	32
50.0	95	5

at least one lesion, ≥ 50% on coronary angiography. Table 9 describes the performance of treadmill testing in these 52 men. The predictive value of 74% was much higher than that found in the asymptomatic, apparently healthy men who only had an abnormal treadmill test.

OTHER SCREENING TECHNIQUES

There are other techniques or tests that could be used to screen asymptomatic men for latent CAD. These include risk-factor screening (1–5), systolic time intervals (47), ballistocardiography (48), resting routine ECG's (49), high-frequency ECG's (50), apexcardiography (42), and kinetocardiography (51). We are now in the early stages of developing a multiple logistic-probability equation using the Duncan-Walker technique (52) based on the results of coronary angiography in 325 aircrewmen (46). This should enable the prediction of the probability of having angiographic CAD in USAF aircrewmen. The cardinal risk factors will initially be used, but later additional data, such as the resting ECG, functional performance, and arrhythmia prevalence, will be used. We hope that this will improve the sensitivity and specificity of screening for latent CAD in our population.

Annual resting ECG's hold great promise as a screening tool for CAD. However, the sensitivity and specificity limitations of resting ECG's are already apparent (46, 53). ECG abnormalities do not have as high a predictive value for CAD in asymptomatic populations as they do in hospital or clinic patients (46) because of the differences in prevalence. Follow-up studies will have to be performed to actually determine these values as well as the cost effectiveness of this screening technique. The addition of computer techniques to the ECG can certainly facilitate ECG testing procedures as well as enable measurements to be made more accurately. However, the impact of computer techniques upon sensitivity and specificity are uncertain and need to be demonstrated. At this time, the high cost of programming may outweigh the advantages of computers, but the development of microprocessors may revolutionize this situation.

Table 9 Analysis of how maximal treadmill testing (MTT) functioned in 52 patients studied with coronary angiography because of either a probable myocardial infarct or mild angina pectoris.

Percentage with abnormal MTT test	Percentage of those with abnormal MTT who had significant angiographic CAD (predictive value)	False-positive rate	Percentage of those with significant angiographic CAD who had abnormal MTT (sensitivity)	Relative risk of having angiographic CAD if MTT response is abnormal vs normal
65 $\left(\frac{34}{52}\right)$	74 $\left(\frac{25}{34}\right)$	26 $\left(\frac{9}{34}\right)$	78 $\left(\frac{25}{32}\right)$	1.9X

RELATIONSHIP OF SENSITIVITY AND SPECIFICITY

No matter what techniques are used, there is a trade-off between sensitivity and specificity. The more specific a particular test is (i.e. the more able it is to determine who is disease-free), the less sensitive it is. This trade-off is altered by adjusting the criteria for abnormal according to the application of the test. When the criteria for an abnormal ECG response to exercise testing are altered to make it more specific for CAD, sensitivity has lessened (10). Certainly, when screening aviators who have jobs where sudden incapacitation can compromise public safety, sensitivity is of utmost importance.

In men less than 30 years of age, the prevalence of significant CAD is so low that tests screening for electrical or mechanical signs of dysfunction are inappropriate (7, 8). The effective yield of such tests is so low that it is better to screen for factors that appear to cause the atherosclerotic process. Such risk-factor screening should be performed in children and young people in order to prevent or stop CAD in its earliest stages (54).

The use of additional data gathered from treadmill testing has been stressed by other investigators. These data include maximal heart rate and blood pressure, functional aerobic capacity, and the heart-rate and blood-pressure response to submaximal work loads (38, 55–57). However, the criteria for abnormality and their effect on the sensitivity and specificity of these measurements have not been adequately determined (58).

SUMMARY

In summary, near-maximal or maximal exercise testing has a sensitivity of approximately 60% and a specificity of approximately 90% for coronary atherosclerotic heart disease. When screening asymptomatic men with exercise testing, an abnormal response identifies a group of men at very high risk for coronary artery disease.

However, the predictive value limitations are obvious and the false-positive problem must be realized. At present, there is no second line of noninvasive studies that can separate an exercise-test false positive from a true positive with certainty. Risk-factor consideration may help separate them. The sensitivity limitations of exercise testing must be especially considered when evaluating people at high risk for CAD.

An abnormal test response does not absolutely predict the presence of CAD and a normal response does not rule out its possibility. In appropriate instances where coronary angiography can be performed at minimal risk and when it is justified for reasons of public safety or individual well-being, this procedure can give a reasonably definitive answer. Creation of iatrogenic "cardiac cripples" can be the most common complication of screening tests and should be avoided. Therefore, good clinical judgment needs to be used in conjunction with any screening test.

Literature Cited

1. Intersociety Commission for Heart Disease 1970. Primary prevention of the atherosclerotic diseases. *Circulation* 42A:55–67
2. Stamler, J. 1971. Acute myocardial infarction—progress in primary prevention. *Br. Heart J.* 33 (Suppl):145–53
3. Froelicher, V. F. 1973. The dietary prevention of atherosclerosis. *Am. Fam. Physician* 7:79–85
4. Kannel, W. B. 1976. Some lessons in cardiovascular epidemiology from Framingham. *Am. J. Cardiol.* 37:269–77
5. Keys, A. 1975. Coronary heart disease —the global picture. *Atherosclerosis* 22:149–92
6. Vecchio, T. J. 1966. Predictive value of a single diagnostic test in unselected population. *N. Engl. J. Med.* 274: 1171–74
7. Froelicher, V. F. 1973. The application of electrocardiographic screening and exercise testing to preventive cardiology. *Prev. Med.* 2:592–99
8. Froelicher, V. F., Thomas, M., Pillow, C., Lancaster, M. C. 1974. An epidemiological study of asymptomatic men screened by maximal treadmill testing for latent coronary artery disease. *Am. J. Cardiol.* 34:770–75
9. Froelicher, V. F., Thompson, A. J., Yanowitz, F., Lancaster, M. C. 1975. Treadmill exercise testing at the USAF-SAM: physiological responses in aircrewmen and the detection of latent coronary artery disease. *AGARDOGRAPH No. 210.* NASA, Langley Field, Va.
10. Froelicher, V. F., Thompson, A. J., Longo, M. R., Triebwasser, J. H., Lancaster, M. C. 1976. The value of exercise testing for screening asymptomatic men for latent coronary artery disease. *Prog. Cardiovasc. Dis.* 18:265–72
11. Friesinger, G., Biern, R., Likar, I., Mason, R. 1972. Exercise ECG and vasoregulatory abnormalities. *Am. J. Cardiol.* 30:733–38
12. Lary, D., Goldschlager, N. 1974. Electrocardiographic changes during hyperventilation resembling myocardial ischemia in patients with normal coronary arteriograms. *Am. Heart J.* 87: 383–90
13. Simonson, E., Keys, A. 1950. The effect of an ordinary meal on the electrocardiogram. *Circulation* 1:1000–5
14. Bruce, R. A. 1968. The effects of digoxin on fatiguing static and dynamic exercise in man. *Clin. Sci.* 34:29–33
15. Whinnery, J., Froelicher, V. F. et al 1976. The electrocardiographic response to maximal treadmill exercise of asymptomatic men with bundle branch block. *Aviat. Space Environ. Med.* 47:1217–25
16. Harris, C., Aronow, W., Parker, D., Kaplan, M. 1973. Treadmill stress test in left ventricular hypertrophy. *Chest* 63:353–57
17. Hellerstein, H., Prozan, G., Liebow, I., Doan, A., Henderson, J. 1961. Two-step exercise test as a test of cardiac function in chronic rheumatic heart disease and in arteriosclerotic heart disease with old myocardial infarction. *Am. J. Cardiol.* 7:234–42
18. Bruce, R. A. 1974. Exercise electrocardiography. In *The Heart,* ed. J. W. Hurst. New York:McGraw-Hill. 3rd ed.
19. Barnard, R., MacAlpin, R., Kattus, A., Buckberg, G. 1973. Ischemic response

to sudden strenuous exercise in healthy men. *Circulation* 48:936–42

20. Berson, A. S., Pipberger, H. V. 1967. Electrocardiographic distortions caused by inadequate high frequency response of direct-writing electrocardiographs. *Am. Heart J.* 74:208–12
21. Froelicher, V. F., Wolthuis, R., Keiser, N., Stewart, A., Fischer, J., Longo, M. R., Triebwasser, J. H., Lancaster, M. C. 1976. Two bipolar exercise ECG leads compared to V_5. *Chest.* 70:611–16
22. Blackburn, H. et al 1968. The exercise electrocardiogram: differences in interpretation. *Am. J. Cardiol.* 21:871–82
23. Schwartz, J. N., Kong, Y., Hackel, D. B., Bartel, A. G. 1975. Comparison of angiographic and postmortem findings in patients with coronary artery disease. *Am. J. Cardiol.* 36:174–78
24. Barresi, V., Manchester, J. H., Shelburne, J. C. 1973. Angiographic evidence of flush coronary occlusion. *J. Am. Med. Assoc.* 225:1111–12
25. Adams, D. F., Abrams, H. L., Ruttley, M. 1972. The roentgen pathology of coronary artery disease. *Semin. Roentgenol.* 7:319–51
26. Arani, D. T., Bunnell, I. L., Greene, D. G. 1975. Lordotic right posterior oblique projection of the left coronary artery. *Circulation* 52:504–6
27. Lintgen, A. B. 1976. Death from myocardial infarction after exercise test with normal result. *J. Am. Med. Assoc.* 235:837–39
28. Paterson, J. C. 1961. Stress, intimal hemorrhage, and coronary occlusion. *J. Occup. Med.* 3:59–63
29. Maseri, A., Mimmo, R., Chierchia, S. 1975. Coronary artery spasm as a cause of acute myocardial ischemia in man. *Chest* 68:625–29
30. Eliot, R. S., Baroldi, G., Leone, A. 1974. Necropsy studies in myocardial infarction with minimal or no coronary luminal reduction due to atherosclerosis. *Circulation* 49:1127–32
31. Richardson, P. J., Livesley, B., Oram, S. 1974. Angina pectoris with normal coronary arteries: transvenous myocardial biopsy in diagnosis. *Lancet* 2:677–80
32. Khan, A. H., Haywood, L. J. 1974. Myocardial infarction in nine patients with radiologically patent coronary arteries. *N. Engl. J. Med.* 291:427–31
33. Mason, R. E., Likar, I., Biern, R. O., Ross, R. S. 1967. Multiple lead exercise electrocardiography. Experience in 107 normal subjects and 67 patients with angina pectoris, and comparison with coronary cine-arteriography in 84 patients. *Circulation* 36:517–22
34. Kasselbaum, D. G., Sutherland, K. I., Judkins, M. P. 1969. A comparison of hypoxemia and exercise electrocardiography in CAD. *Am. Heart J.* 75:759–74
35. Roitman, D., Jones, W. B., Sheffield, L. T. 1970. Comparison of submaximal exercise ECG test with coronary cineangiocardiogram. *Ann. Intern. Med.* 72:641–47
36. Ascoop, C. A., Simoons, M. L., Egmond, W. G., Bruschke, A. V. G. 1971. Exercise test, history, and serum lipid levels in patients with chest pain and normal electrocardiogram at rest: comparison to findings at coronary arteriography. *Am. Heart J.* 82:609–17
37. McHenry, P. L., Phillips, J. F., Knoebel, S. B. 1972. Correlation of computer-quantitated treadmill exercise electrocardiogram with arteriographic location of coronary artery disease. *Am. J. Cardiol.* 30:747–52
38. Bartel, A. G., Behar, V. S., Peter, R. H., Orgain, E. S., Kong, Y. 1974. Graded exercise stress tests in angiographically documented coronary artery disease. *Circulation* 49:348–56
39. Borer, J. S., Brensike, J. F., Redwood, D. R., Epstein, S. E. 1975. Limitations of the electrocardiographic response to exercise in predicting coronary artery disease. *N. Engl. J. Med.* 293:367–72
40. Bruce, R. A., Hornstein, T., Blackman, J. 1968. Myocardial infarction after normal response to exercise. *Circulation* 38:552–55
41. Bruce, R. A., McDonough, J. R. 1969. Stress testing in screening for cardiovascular disease. *Bull. NY Acad. Med.* 45:1288–95
42. Aronow, W. S., Cassidy, J. 1975. Five-year follow-up of double Master's test, maximal treadmill stress test, and resting and postexercise apexcardiogram in asymptomatic persons. *Circulation* 52: 616–18
43. Cumming, G. R., Samm, J., Borysyk, L., Kich, L. 1975. Electrocardiographic changes during exercise in asymptomatic men: 3-year follow-up. *Can. Med. Assoc. J.* 112:578–81
44. Jelliffe, R. W. 1973. Quantitative aspects of clinical judgment. *Am. J. Med.* 55:431–35
45. Galen, R. S., Gambino, S. R. 1975. *Beyond Normality.* New York: Wiley
46. Froelicher, V. F., Thompson, A. J., Wolthuis, R., Fuchs, R., Balusek, R., Longo, M. R., Triebwasser, J. H., Lan-

caster, M. C. 1976. Cardiac catheterization of asymptomatic aircrewmen with ECG abnormalities. *Am. J. Cardiol.* In press

47. Montoye, H. J., Willis, P. W. III, Howard, G. E., Keller, J. B. 1970. Systolic preejection period: patients with heart disease compared to normal subjects. *Arch. Environ. Health* 21:425–31
48. Lynn, T. N., Wolf, S. 1974. The prognostic significance of the ballistocardiogram in ischemic heart disease. *Am. Heart J.* 88:277–80
49. Blackburn, H. 1973. Importance of the electrocardiogram in populations outside the hospital, *Can. Med. Assoc. J.* 108:1262–65
50. Anderson, G. J., Blieden, M. F. 1975. The high frequency electrocardiogram in coronary artery disease. *Am. Heart J.* 89:349–58
51. Eddleman, E. E. 1974. Examination of precordial movements. In *The Heart,* ed. J. W. Hurst, p. 189. New York:McGraw-Hill. 3rd ed.
52. Walker, S. H., Duncan, D. B. 1967. Estimation of probability of an event as a function of several independent variables. *Biometrika* 54:167–79
53. Miller, R. R., Massumi, R. A., Mason, D. T. 1975. The resting electrocardiogram in assessing ischemic heart disease. *Geriatrics* 30:51–57
54. Neufeld, H. N. 1974. Precursors of coronary arteriosclerosis in the pediatric and young adult age groups. *Mod. Concepts Cardiovasc. Dis.* 43:93–98
55. Bruce, R. A. 1974. Editorial: values and limitations of exercise electrocardiography. *Circulation* 50:1–4
56. Bruce, R. A., Fisher, L. D., Cooper, M. N., Gey, G. O. 1974. Separation of effects of cardiovascular disease and age on ventricular function with maximal exercise. *Am. J. Cardiol.* 34:757–63
57. Ellestad, M. H., Wan, M. K. C. 1975. Predictive implications of stress testing. *Circulation* 51:363–69
58. Wolthuis, R., Froelicher, V. F., Fischer, J. 1977. The physiological response of apparently healthy men to treadmill exercise. *Circulation.* In press

Ann. Rev. Med. 1977. 28:13–20
Copyright © 1977 by Annual Reviews Inc. All rights reserved

IONIC BASIS OF MYOCARDIAL CONTRACTILITY[1]

❖7218

Glenn A. Langer, M.D.
The Departments of Medicine and Physiology and the American Heart Association,
Greater Los Angeles Affiliate Cardiovascular Research Laboratory, University of
California at Los Angeles, Center for the Health Sciences, Los Angeles, California 90024

The subject of this chapter is a broad one and it is necessary to be selective in its discussion. I focus on two major areas: (*a*) the control and interrelationships of ionic movements in the myocardial cell and (*b*) the source of the calcium (Ca) that regulates the cell's force development. The first area I present with particular attention to the action of the digitalis glycosides, since these drugs are of considerable clinical interest and because they have been of great aid in the definition of ionic movements. The second area is one of rapid development and is of particular interest to those concerned with preservation of the myocardium under various pathological conditions. There remains a good deal of controversy in both areas, but this is to be expected when our experimental tools are as yet incapable of testing a variety of internally consistent models.

CONTROL OF IONIC MOVEMENTS—THE ACTION OF DIGITALIS

The digitalis compounds are very specific with respect to their action on a variety of cell types. There is no evidence indicating that digitalis does anything but inhibit the Na-K pump at the cellular membrane (1). Schatzmann (2) was the first to show that the drug acts directly upon the pump, followed by the demonstration that radioactively labeled digitalis compounds bind specifically to Na-K ATPase, the enzyme that controls the activity of the pump (3). It would be predicted that inhibition of the pump would result in a net loss of K and a net gain of Na by the affected cell; this has been conclusively demonstrated (4, 5).

Though there is agreement that the action of digitalis is to specifically inhibit Na-K pumping, there remains disagreement among enzymatic studies as to whether

---

[1] Supported by Grant HL 11351–10 from the US Public Health Service and by a Grant from the Castera Foundation.

this inhibition can or cannot be dissociated from the positive effect of the drug on cardiac contractility. Okita et al (6), using short-acting compounds and drug washout experiments in isolated hearts, has demonstrated that although the Na-K ATPase isolated from such preparations is still significantly inhibited, the positive inotropic effect was no longer evident. These studies are to be contrasted with those of Allen et al (7) and Ku et al (8), in which there was a positive correlation between enzyme inhibition and increased contractile force. It is possible that the reason for these different results lies in the need to compare the functional state of the tissue in vivo with the state of the enzyme in vitro after fractionation. In the in vivo state, the enzyme demonstrates an asymmetric response to Na and K at the inside and outside of the membrane and the ionic requirements are quite specific. It is not surprising, therefore, that relatively small differences in fractionation technique and in the suspension medium would affect enzymatic activity.

Using intact, functional ventricular muscle, it has not been possible to dissociate the production of positive inotropy by digitalis compounds from Na-K pump inhibition (9–12). The inhibition need not result in very large ionic changes in the cell, e.g. maximal inotropy is coincident with an increase of 2–3 mmole Na/kg wet wt of ventricle. This represents a gain of 15–20% in total cellular Na (13, 14), but it should be emphasized that this relatively small gain could produce a very large increase in Na concentration in a restricted or compartmentalized portion of the cell. It is clearly established that approximately 80% of ventricular Na is bound or compartmentalized and not homogeneously distributed in the cell water (15).

I believe, then, that the evidence available at present favors the concept that by some mechanism an inhibition of the Na-K pump leads to augmented contractile force. If this is the case, it raises interesting possibilities as to the basic mechanisms of ionic control of contractility. How might the glycoside-induced net gain in cellular Na and/or net loss in cellular K result in a more forceful contraction?

Before I examine some of the models proposed, it is necessary to note the effect of digitalis on Ca movements at and across the cellular membrane. It is well documented that Ca influx is augmented after glycoside administration (5, 9, 16–18) and it can reasonably be concluded that it is an increase of Ca at the myofilaments that serves as the final mediator of the positive inotropism. This, then, raises the possibility that the Na and K movements summarized above could in some way be related to the Ca influx.

The possibility of a system in which Na and Ca movements are related was discussed in 1964 by Repke (19) and by Langer (20). Though a Na-Ca flux relationship was supported by further experiments in heart tissue (21), it remained for experiments on the perfused squid axon (22) to suggest a more detailed model. In this preparation it was clearly demonstrated that upon an increase in Na concentration at the inner side of the membrane, Ca influx was significantly augmented. Indirect evidence for the same sequence has now been found in heart muscle (14). These demonstrations provide a plausible link between the primary action of the digitalis compounds, i.e. Na-K pump inhibition with resultant intracellular Na accumulation, and the increased Ca influx believed to be directly responsible for the augmented force development.

Despite the clear evidence for enhanced influx of Ca with glycoside administration, there is no evidence that this movement is associated with an increased inward current that might be expected to be manifest during the excitation-contraction coupling period. Two recent studies (23, 24) using voltage-clamp technique could find no change in the amplitude of the so-called slow inward current, which is believed to be associated with electrogenic Ca movements into the cell. This was despite the administration of markedly inotropic concentrations of digitalis. This means that if the Ca flux that is responsible for the force development moves across the sarcolemmal membrane (see below), it moves in an electrically neutral manner. Such movement would be consistent with a model that links outward Na movement to inward Ca movement via a carrier moving across the cardiac sarcolemma (25) —much as has been demonstrated for the squid axon membrane.

It has been proposed (24) that the augmented inward Ca movement is linked not to Na efflux, but directly to the augmented K efflux that digitalis produces (see above). Such K-Ca coupling could serve to explain the electroneutrality of the system as well as a Na-Ca coupling. It is, however, unlikely that a K-Ca coupling plays a role. It has been shown that both respiratory acidosis (26) and magnesium (12) are capable of a specific blockade of K efflux from the heart, and both interventions have been used to eliminate the increased "K leak" induced by the glycosides (12, 27). A marked inotropic effect occurred without an increase of K efflux. K-uptake experiments did, however, clearly demonstrate that the Na-K pump was inhibited and therefore a net gain in cellular Na was induced. This is consistent with the proposed Na-Ca exchange model.

Further support for the proposal that digitalis exerts its inotropic effect by augmentation of a Na-Ca carrier system comes from our recent study on the neonatal rat heart, which demonstrates a striking decrease in its sensitivity to ouabain over the first 3 weeks of life (10). During the first week after birth, 5×10^{-5} M ouabain produces a 140% increase in force development. The same dose of ouabain produces only a 12% increase in force in hearts removed from 3-week-old rats. During the two-week interval over which resistance to ouabain is developing, the duration of the plateau phase of the ventricular action potential decreases markedly. This decrease was due, in part, to closure of a "slow channel" for Na influx, which would have the effect of diminishing the Na influx associated with excitation. It would follow that as the rat ventricle aged, the Na influx per excitation would decrease and, therefore, there would be less increase in cellular Na when the Na-K pump was inhibited. The development of insensitivity to glycoside as the rat ages is, then, consistent with less accumulation of cellular Na upon administration of the drug. Additional support for this sequence is the demonstration that the markedly positive inotropic response to ouabain exhibited by the one-week-old heart was eliminated when these hearts were perfused with a solution containing 50% $[Na]_o$ and 25% $[Ca]_o$ as compared with the normal perfusate. This solution does not alter control-force development since the ratio $[Ca]_o/[Na]_o^2$ is held constant and therefore the "availability" of Ca for excitation-contraction (E-C) coupling is unchanged (28). The reduction in $[Na]_o$ does, however, diminish the Na influx per excitation and is then comparable to the aging process.

The foregoing discussion has used the action of digitalis to review some of the present concepts of the ionic control of myocardial contractility. It has been presented with a bias toward the importance of a Na-Ca exchange system and it should be emphasized that others are of the opinion that K efflux cannot be ruled out as an important mediator of inotropy (24), or that digitalis causes a specific conformational change in the Na-K ATPase, which then allosterically affects adjacent membrane phospholipids to increase their affinity for Ca (29). It is proposed that this Ca is released directly to the sarcomeres without the need for augmentation of a Na-Ca exchange system. The subject obviously requires further definition before a final model can be assumed. In addition, implicit in the Na-Ca change model as I have reviewed it is the concept that the major source of contractile-dependent Ca in the heart is derived from sites in rapid equilibrium with the interstitial space, i.e. not derived from the lateral cisternal components of the sarcotubular system as is presently accepted for skeletal muscle (30). The source of Ca is another controversial area currently under investigation in many laboratories and deserves a brief review.

SOURCE OF CONTRACTILE-DEPENDENT CALCIUM

Solaro et al (31) have shown that approximately 35 μmoles Ca/kg wet wt of heart muscle delivered to the myofilaments is sufficient to produce 90% maximum force development. This value is to be compared to the amount of Ca calculated to cross the sarcolemmal membrane using voltage-clamp data with integration of the slow current (believed to represent Ca influx) with respect to time. When such integration is done, the *maximum* Ca influx associated with full mechanical activation is between 5 and 10 μmole Ca influx $(kg)^{-1}$ $(beat)^{-1}$ (32, 33), with most values an order of magnitude less. Solaro et al (31) indicated that 8 μmole/kg wet wt were required for 10% maximum force development; thus the value for Ca influx derived from analysis of its electrogenic movement is barely enough to exceed mechanical threshold, much less account for full contractile force. Therefore, there must be considerable Ca flowing to the myofilaments during activation that is not electrogenic.

There are two general models currently proposed to account for the additional Ca required for the major portion of activation. One of the models places the Ca in an intracellular storage site (probably the lateral cisternae of the sarcotubular system) to be released by a potential change at the sarcotubular membrane or by a small amount of Ca coming across the sarcolemma (so-called triggered or regenerative Ca release) (34, 35). Except for the absolute quantities involved, this model is similar to that now generally accepted for skeletal muscle, where the source of coupling Ca is the intracellular sarcotubular system (30). Since the major portion of the Ca is derived intracellularly, it would not contribute to the slow inward current, and the model is therefore consistent with voltage-clamp findings.

The model is not, however, consistent with a number of findings that indicate major differences between heart and skeletal muscle with respect to the E-C coupling sequence (36):

1. Perfusion with solution that contains no Ca causes heart-muscle force to decline with a half-time of 15–60 sec, whereas skeletal muscle force declines with a half-time of 30 min or longer.
2. Various cations, such as lanthanum, when added to heart muscle perfusate, are capable of producing complete blockade of E-C coupling, but are almost totally ineffective in skeletal muscle tissue.
3. Induction of respiratory acidosis results in a marked reduction of force in heart muscle and produces no or minimal effects in skeletal muscle.
4. Correlation of the pattern of decline of force with the pattern of Ca exchange (using ^{45}Ca tracer) indicates that E-C uncoupling in the heart occurs as Ca is lost from a very rapidly exchangeable pool ($t_{1/2} = 1$ min or less). Similar studies on skeletal muscle indicate that the rapidly exchangeable Ca pool plays little role in the support of force.
5. Recent evidence indicates that "triggered" Ca release is not physiological in skeletal muscle. It remains controversial as to whether or not the process plays a role in normal heart muscle. There is evidence that it is operative in the skinned cells from rat ventricle (37), but is more difficult to demonstrate in skinned cells from ventricles of other mammalian species (38).

If triggered release plays a physiological role in the E-C coupling sequence in the heart, a model in which Ca is derived from an intracellular source is certainly possible. A conclusive answer will probably not be obtained until a cardiac single-fiber preparation with intact (unskinned) membrane is achieved. As the preceding points indicate, however, a model that places the coupling Ca at the cellular surface is a very real alternative.

If the Ca is derived from the cellular surface, it has to be bound or stored in association with some structure. External to the unit membrane (approximately 90 Å thick) there is a structure usually referred to as the basement membrane. This is a misnomer. The structure is composed of an inner layer of glycoprotein (approximately 200 Å thick) called the cell coat and an outer layer (approximately 400 Å thick) termed the external lamina. These layers contain, among other things, large amounts of acidic mucopolysaccharides, which are therefore anionic at the pH of the extracellular space (39). It is reasonable to expect that this area of the cell would be capable of binding significant amounts of Ca, and evidence for this has been presented (40). It can be visualized that the Ca "stored" within the surface layers would supply two general systems—one a system of pores or conductive channels through the unit membrane and the other the carrier system capable of moving Na outward and Ca inward. The pores are visualized to function as selective channels for the electrogenic movement of ions (measured by voltage-clamp technique), including Ca. The carrier system would transport ions in an electroneutral manner as previously described.

The surface layers (surface coat and external lamina) and their possible role in heart-muscle function have been virtually neglected. The major importance of surface constituents is emphasized by our recent study on sialic acid (41). The sialic acids occupy peripheral, terminal positions in the oligosaccharide portions of the glycoproteins and glycolipids that make up a large component of cellular surface-coat material (42) and they have a particular affinity for Ca (43). We have found that removal of sialic acid from cultured heart cells with highly purified neuraminidase (specific for cleaving sialic acid) markedly increases cellular Ca exchangeability (41). It is striking that as the membrane becomes nonselective with respect to Ca permeability, its selectivity for K is maintained.

These results indicate an important role for the surface coat/external lamina complex in the regulation of Ca exchange in the heart. Work in progress in our laboratory strongly indicates that the integrity of the surface layers plays a significant role in the determination of reversibility after myocardial cell injury such as follows an ischemic period. It is likely that Ca overloading of the cell is a major contributor to progressive cellular damage during reperfusion (44). Though it is possible that failure of Ca-sequestration mechanisms plays a role in the overloading, it seems very likely that increased influx also occurs. If the surface layers, and more specifically sialic acid, play a critical role in the control of Ca entry, they deserve increased attention in our studies of ionic exchange in both normal and pathological states.

CONCLUSION

Knowledge about the basic physiology of mammalian cardiac muscle has lagged behind that of skeletal muscle because the tissue is more difficult to deal with experimentally. The cells are much smaller, they do not tolerate anaerobic conditions very well, and they cannot be tetanized. Therefore as knowledge has developed about skeletal muscle, there has been a tendency to extrapolate to the myocardium. In many instances such extrapolation has been misleading, especially in the area of ionic control of contractility. The mammalian heart, with its "all or none" contraction, must have an intrinsic mode of modulating its force development. There is no mechanism as in skeletal muscle for recruitment of more or less motor units as force or work required varies. The heart cell uses two basic mechanisms in the regulation of its contraction—the variation of force with initial cellular length (the classic Frank-Starling mechanism) and the variation of the amount of Ca that reaches the myofilaments. This brief review has focused on the second mechanism.

I have emphasized the following points:

1. The source of contractile-dependent Ca in the heart is at the cell surface in rapid equilibrium with the extracellular space.
2. It is likely that a large portion of this surface Ca is bound to negatively charged molecules located in the surface coat/external lamina complex. Sialic acid is a major constituent of the surface complex and recent studies indicate that it is important in the control of Ca permeability in the myocardial cell.
3. The Ca derived from the surface complex crosses the sarcolemmal unit membrane by two routes: (*a*) a "pore" system: passage through this system is electrogenic and measurable by voltage-clamp technique; (*b*) a "carrier system:" passage through this system is coupled to the outward movement of another

cation—most likely Na—such that it is electroneutral and therefore, the inward Ca movements are not detectable by voltage-clamp technique.

4. The primary action of the digitalis glycosides is to inhibit the Na-K pump. This increases Na intracellularly, which in turn increases the Na-Ca carrier activity with the result that more Ca is delivered to the myofilaments. This produces the well-known positive inotropic effect.
5. Finally, there remains considerable controversy with respect to the foregoing points. At this time, they should be viewed as components of a working model that awaits the development of experimental tools to test further its validity.

Literature Cited

1. Glynn, I. M. 1964. The action of cardiac glycosides on ion movements. *Pharmacol. Rev.* 16:381–407
2. Schatzmann, H. J. 1953. Herzglykoside als Hemmstaffe für den activen Kalium- und Natriumtransport durch die Erythrocytenmembran. *Helv. Physiol. Pharmacol. Acta* 11:346–54
3. Matsui, H., Schwartz, A. 1968. Mechanism of cardiac glycoside inhibition of the (Na^+-K^+)-dependent ATPase from cardiac tissue. *Biochim. Biophys. Acta* 151:655–63
4. Langer, G. A. 1972. Effects of digitalis on myocardial ionic exchange. *Circulation* 46:180–7
5. Lee, K. S., Klaus, W. 1971. The subcellular basis for the mechanism of inotropic action of the cardiac glycosides. *Pharmacology* 23:193–261
6. Okita, G. T., Ten Eick, R. E., Richardson, F. 1974. Inhibition of Na^+, K^+ ATPase and digitalis action: dissociation from inotropic effects and its role in digitalis cardiotoxicity. In *Properties and Functions of (Na^+-K^+)-Activated Adenosinetriphosphatase*, ed. A. Askari, 242:658–70. *Ann. NY Acad. Sci.*
7. Allen, J. C., Entman, M. L., Schwartz, A. 1975. The nature of the transport adenosine triphosphatase-digitalis complex. VIII. The relationship between in vivo formed (^{3}H-ouabain-Na^+-K^+-adenosine triphosphatase) complex and ouabain induced positive inotropism. *J. Pharm. Exp. Ther.* 192:105–12
8. Ku, D., Akera, T., Pew, C. L., Brody, T. M. 1974. Cardiac glycosides: correlates among Na^+, K^+-ATPase, sodium pump and contractility in guinea pig heart. *Naunyn. Schmied. Arch. Pharmakol.* 285:185–200
9. Langer, G. A., Serena, S. D. 1970. Effects of strophanthidin upon contraction and ionic exchange in ventricular myocardium: relation to control of active state. *J. Mol. Cell. Cardiol.* 1:65–90
10. Langer, G. A., Brady, A. J., Tan, S. T., Serena, S. D. 1975. Correlation of the glycoside response, the force staircase and the action potential configuration in the neonatal rat heart. *Circ. Res.* 34: 744–52
11. Poole-Wilson, P. A., Langer, G. A. 1975. Glycoside inotrophy in the absence of an increase in potassium efflux in the rabbit heart. *Circ. Res.* 37:390–5
12. Shine, K. I., Douglas, A. M. 1975. Magnesium effects in rabbit ventricle. *Am. J. Physiol.* 228:1545–54
13. Page, E., Page, E. G. 1968. Distribution of ions and water between tissue compartments in the perfused left ventricle of the rat heart. *Circ. Res.* 22: 435–46
14. Tillisch, J., Langer, G. A. 1974. Myocardial mechanical responses and ionic exchange in high sodium perfusate. *Circ. Res.* 34:40–50
15. Lee, C. O., Fozzard, H. A. 1975. Activities of potassium and sodium ions in rabbit heart muscle. *J. Gen. Physiol.* 65:695–708
16. Bailey, L. E., Sures, H. A. 1971. The effect of ouabain on the washout and uptake of calcium in the isolated cat heart. *J. Pharm. Exp. Ther.* 178:259–70
17. Nayler, W. G. 1973. An effect of ouabain on the superficially-located stores of calcium in cardiac muscle cells. *J. Mol. Cell. Cardiol.* 5:101–10
18. Langer, G. A. 1973. Heart: Excitation-contraction coupling. *Ann. Rev. Physiol.* 35:55–86
19. Repke, K. 1964. Über den biochemischen Wirkingsmodus von Digitalis. *Klin. Wochenschr.* 41:157–65
20. Langer, G. A. 1964. Kinetic studies of

calcium distribution in ventricular muscle of the dog. *Circ. Res.* 15:393–405

21. Langer, G. A. 1967. Sodium exchange in dog ventricular muscle. Relation to frequency of contraction and its possible role in the control of myocardial contractility. *J. Gen. Physiol.* 50: 1221–39
22. Baker, P. F., Blaustein, M. P., Hodgkin, A. L., Steinhardt, R. A. 1969. The influence of calcium on sodium efflux in squid axons. *J. Physiol. London* 200:431–58
23. McDonald, T. F., Nawrath, H., Trautwein, W. 1975. Membrane currents and tension in cat ventricular muscle treated with cardiac glycosides. *Circ. Res.* 37:674–82
24. Greenspan, A. M., Morad, M. 1975. Electromechanical studies on the inotropic effects of acetylstrophanthidin in ventricular muscle. *J. Physiol. London* 253:357–84
25. Langer, G. A. 1971. The intrinsic control of myocardial contraction–ionic factors. *N. Engl. J. Med.* 285:1065–71
26. Poole-Wilson, P. A., Langer, G. A. 1975. Effect of pH on ionic exchange and function in rat and rabbit myocardium. *Am. J. Physiol.* 229:570–81
27. Poole-Wilson, P. A., Langer, G. A. 1975. Glycoside inotropy in the absence of an increase in potassium efflux in the rabbit heart. *Circ. Res.* 37:390–95
28. Luttgau, H. C., Niedergerke, R. 1958. The antagonism between Ca and Na ions on the frog's heart. *J. Physiol. London* 143:486–505
29. Gervais, A., Lindenmayer, G. E., Entman, M. L., Pitts, B. J. R., Lane, L. K., Schwartz, A. 1975. A mechanism for ouabain action involving calcium-Na^+, K^+-ATPase interaction. *Circulation* 52: II, 25
30. Winegrad, S. 1968. Intracellular calcium movements of frog skeletal muscle during recovery from tetanus. *J. Gen. Physiol.* 51:65–83
31. Solaro, R. J., Wise, R. M., Shiner, J. S., Briggs, F. N. 1974. Calcium requirements for cardiac myofibrillar activation. *Circ. Res.* 34:525–30
32. Beeler, G. W., Reuter, H. 1970. The relation between membrane potential, membrane currents and activation of contraction in ventricular myocardial fibers. *J. Physiol. London* 207:211–29
33. New, W., Trautwein, W. 1972. Inward currents in mammalian myocardium. *Pfluegers Arch.* 334:1–23
34. Morad, M., Goldman, Y. 1973. Excitation-contraction coupling in heart muscle: membrane control of development of tension. In *Progress in Biophysics and Molecular Biology,* ed. A. J. V. Butler, D. Noble, 27:257–313. Oxford and New York: Pergamon
35. Reuter, H. 1974. Exchange of calcium ions in the mammalian myocardium. *Circ. Res.* 34:599–605
36. Langer, G. A., Frank, J. S., Brady, A. J. 1976. The myocardium. *Int. Rev. Physiol.* 9:191–237
37. Fabiato, A., Fabiato, F. 1975. Contractions induced by a calcium-triggered release of calcium from the sarcoplasmic reticulum of single skinned cardiac cells. *J. Physiol. London* 249: 469–95
38. Fabiato, A., Fabiato, F. 1973. Activation of skinned cardiac cells. Subcellular effects of cardioactive drugs. *Eur. J. Cardiol.* 1:143–55
39. Martinez-Palomo, A. 1970. The surface coats of animal cells. *Int. Rev. Cytol.* 29:29–75
40. Langer, G. A., Frank, J. S. 1972. Lanthanum in heart cell culture: Effect on calcium exchange correlated with its localization. *J. Cell Biol.* 54:41–55
41. Langer, G. A., Frank, J. S., Nudd, L. M., Seraydarian, K. 1976. Sialic acid: effect of removal on calcium exchangeability of cultured heart cells. *Science.* 193:1013–15
42. Steck, T. L. 1974. The organization of proteins in the human red blood cell membrane. *J. Cell Biol.* 62:1–19
43. Winzler, R. J. 1970. Carbohydrates in cell surfaces. *Int. Rev. Cytol.* 29:77–125
44. Jennings, R. B., Ganote, C. E. 1974. Structural changes in myocardium during acute ischemia. *Circ. Res.* 35 (3): 156–72

Ann. Rev. Med. 1977. 28:21–35
Copyright © 1977 by Annual Reviews Inc. All rights reserved

CORNEAL TRANSPLANTATION ❖7219

S. Lance Forstot, M.D. and Herbert E. Kaufman, M.D.
Department of Ophthalmology, University of Florida, College of Medicine, Gainesville, Florida 32610

INTRODUCTION

The first successful penetrating corneal transplant in man was reported just over 70 years ago (1). It is only during the last 25 years, however, that this type of surgery has achieved a reasonable chance of success. More recently, in the last 5 to 10 years, corneal transplantation has been improved to the point where some transplants have a 95% success rate and most transplants have better than an 80% chance of the cornea remaining clear. Previously hopeless cases, such as chemical burns, have about 50% chance of retaining clarity. The reason for this marked improvement in prognosis for clear corneal transplants appears to rest on three advances: (*a*) improved surgical methodology, technology, and instrumentation; (*b*) a better understanding of corneal biology, especially the recognition of the importance of the corneal endothelium and its preservation; and (*c*) anti-inflammatory agents to better control graft rejection.

Before beginning a more thorough discussion of each of these factors, we describe the two major types of corneal transplant or keratoplasty. The first type is a lamellar keratoplasty, or split-thickness transplant (Figure 1) and the second is a penetrating keratoplasty, or full-thickness transplant (Figure 2). In lamellar keratoplasty a thin partial thickness of corneal tissue is excised from the host eye, leaving a portion of the stroma and the entire endothelium. This is replaced by a matching partial thickness of corneal tissue from the donor eye. Prior to modern technology, lamellar keratoplasty was used extensively because healthy living tissue was less critical, precise surgical apposition less necessary, and its prognosis for visual rehabilitation was greater than that for penetrating keratoplasty, which tended to have greater complications. However, in lamellar keratoplasty the corneal healing process often results in a variable-interface scar between host and donor tissue with less than normal corneal transparency. Although lamellar transplantation is usable for superficial scars and ulcers, it is not suitable for full-thickness stromal or endothelial damage. Recent advances have tremendously improved the prognosis for a clear penetrating keratoplasty, which may afford better visual acuity and which can rectify deep pathology. Therefore, the penetrating keratoplasty has become the

P 24845
~~1800~~
UNIVERSITY COLLEGE OF MEDICINE LIBRARY C. U.

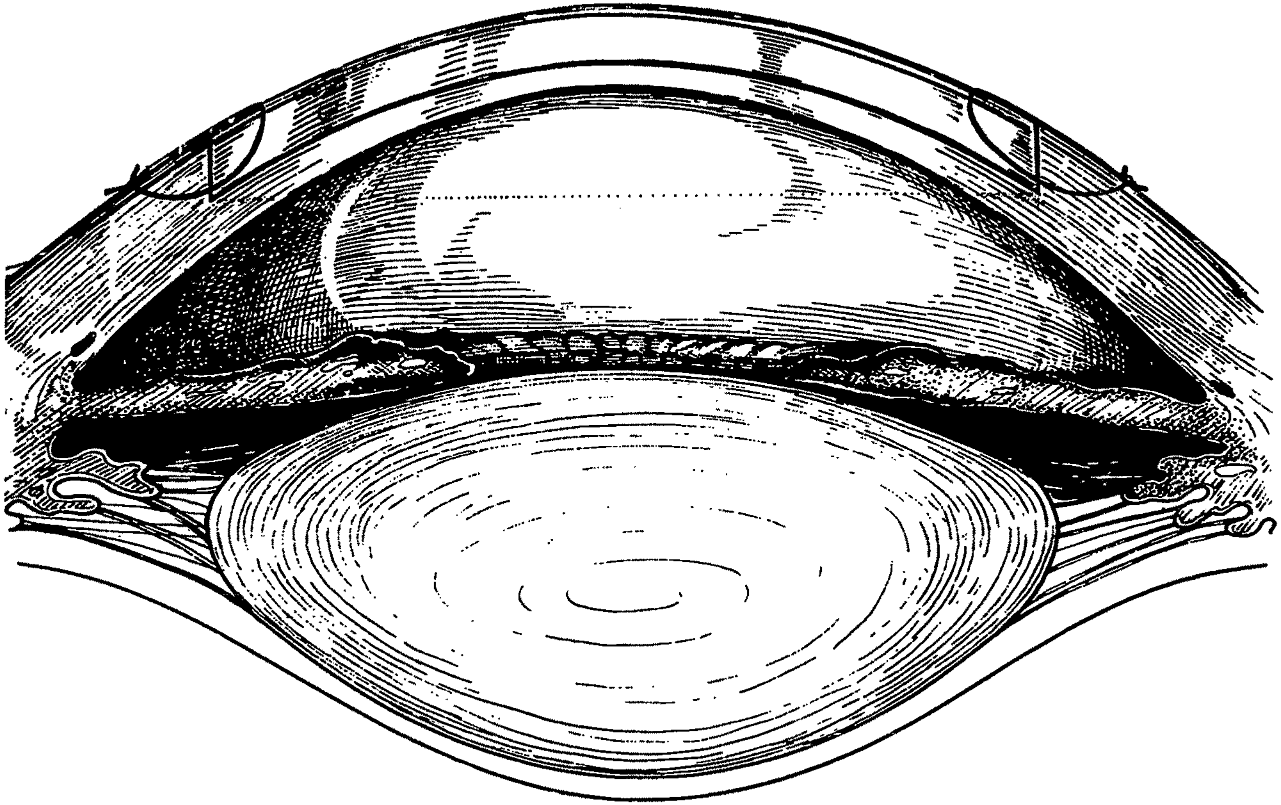


Figure 1 Lamellar transplant, now used mainly for preliminary repair, in which only a thin layer of cornea is replaced. Note that the sutures do not pierce the undersurface of the donor or host cornea of either lamellar or penetrating transplant, a practice necessitated by the extreme sensitivity of the corneal endothelium and made considerably easier by the fineness of modern materials. Reprinted with permission from *Hosp.*

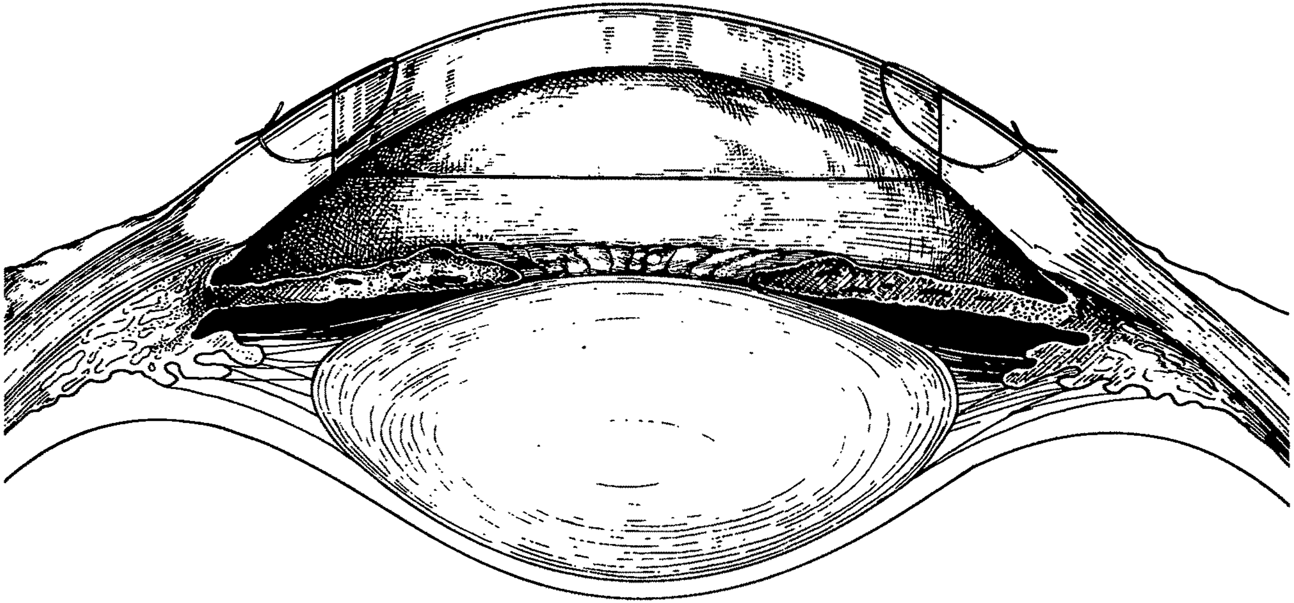


Figure 2 Cross section of a penetrating transplant, in which the full thickness of the cornea is removed and a full-thickness donor button substituted. Reprinted with permission from *Hosp. Pract.* 8:69–75 (1973).

much more common procedure for restoration of vision. Lamellar keratoplasty is now more often used to strengthen penetrating peripheral corneal ulcers where some loss of transparency will not affect visual acuity, or it is used for preliminary repair of small corneal perforations to maintain the integrity of the eye. In addition, lamellar grafts are also used for tectonic buildup of thinned corneas so that a later full-thickness keratoplasty can be performed with a greater probability of success. Lamellar keratoplasties may also be performed as part of "refractive" keratoplasties in patients with unilateral large refractive errors to reduce these by changing the curvature of the corneal disc.

HISTORY

There are several textbooks on corneal transplantation that contain detailed histories (2–5); only a brief history is presented here. In 1789 de Quengsy suggested the possibility of replacing opaque cornea with a transparent material to restore vision. Erasmus Darwin in 1796 suggested cutting out a portion of the opaque cornea to allow healing with formation of a "transparent scar." However, corneal transplantation did not begin until the nineteenth century. Himly (1813) suggested that one might replace opaque cornea in one animal with clear cornea from another. However, it was his pupil, Frans Reisinger, in 1818, who finally went beyond "suggestions" to initiate corneal-transplant experiments in animals. It was Reisinger who suggested the name "keratoplasty," the procedure of corneal transplantation. Although he worked only with animals, he suggested in 1824 that transparent animal cornea might be used to replace opaque human cornea. Though Reisinger's work stimulated great interest, his idea was considered by a prominent contemporary as "one of the most audacious fantasies."

The first human transplant recorded was by Kissam in 1844 when he reported an unsuccessful keratoplasty using the cornea of a pig as donor. Because of the difficulties encountered with the keratoplasty procedure, von Walther recommended replacing only the anterior layers of the cornea and subsequently lamellar transplants attracted much attention. In 1877 von Hippel reported a lamellar keratoplasty in man using dog cornea, but it was also unsuccessful. The lamellar technique of von Hippel, best described in 1888, outlined the use of a trephine and some other features of keratoplasty still used today.

Power in 1872 had suggested that one needed homoplastic tissue for donor material, but his reason for this was related to matching of corneal thickness, not to antigenic differences in species. Although many attempts at penetrating corneal transplants were made, it was not until the twentieth century (1906) that a success was reported. It was Zirm who reported a case performed the previous year, when he had transplanted a cornea from an enucleated eye of an 11-year-old boy into a man with corneal scarring from a lime burn. Following this report, interest in penetrating keratoplasty was again stimulated and the foundations of modern corneal transplantation were laid, especially in the 1920s and 1930s by Elschnig (6) in Prague, and Filatov in Odessa (7). They showed that corneal transplants could be performed with some chance of success. Slow progress was made prior to World

War II with contributions from all parts of the world: Castroviejo (USA), Tudor-Thomas (Great Britain), Nizetic (Yugoslavia), Imre (Hungary), Franceschetti (Switzerland), Arruga (Spain), Friede & Lohlein (Germany), and Vannas (Finland) (2). However, the most rapid progress and greatest strides in keratoplasty followed World War II. The introduction of antibiotics, steroids, better instrumentation, and an understanding of the biology of the cornea and vitreous humor led to the great increase in prognosis for clear keratoplasties recorded in the 1950s and 1960s.

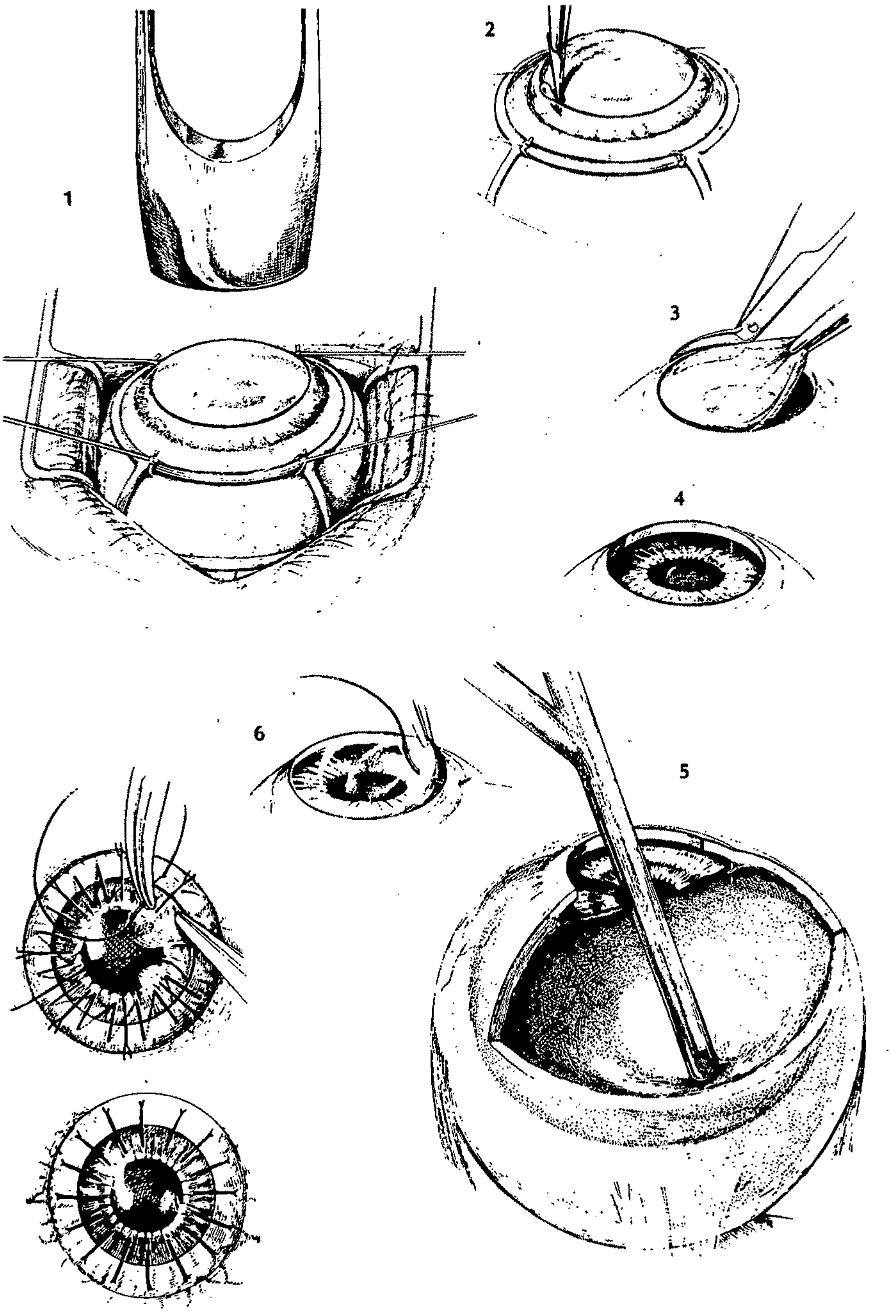
DISCUSSION

It was the mid-twentieth century advances that transformed the nineteenth-century "audacious fantasy" into practical, commonplace reality.

The first significant problem encountered in attempting a keratoplasty was the corneal wound. How could one approximate the corneal tissue to form a watertight wound that would retain aqueous humor within the anterior chamber and maintain a bond, while allowing a period for healing to produce a physiologic union? Early efforts to provide wound integrity were by use of splints over the top of the transplant rather than appositional sutures through the tough, thin (0.5 mm) cornea. These included suturing the lids, conjunctival flaps, disc-like splints placed over the corneal button, and later, overlying sutures. Finally, direct corneal apposition was performed using needles and suture material from donor to host. Because of the delicate manipulations involved, operating loupes of 2X to 4X magnification were required. However, modern technology has provided operating microscopes that now give clear operating fields with 6X to 20X magnification. Because of these developments, newer and finer needles and sutures could be developed. These sutures, although fine, have remarkable tensile strength. Before all these advances, full-thickness grafts were prone to blowout with sudden movement, inadvertent bending, or a sudden blow—even many months following the operation. In the early years, not so long ago, absolute inactivity was a must in the early postoperative period and relative inactivity was required for many months.

The strength of the corneal bond is important, but precision of the bond is as important in preventing opacification of the graft. One needs a watertight junction to prevent the leakage of aqueous humor and development of corneal adhesion to other intraocular tissue (e.g. iris, lens, or vitreous humor), and also distortion of the surface that would give severe astigmatism and optical errors. Poor healing can lead to infection or ultimate opacification, due to graft rejection. The cornea is only 0.5 mm thick. The commonly used 10-0 nylon suture is 23 μm in diameter, requiring surgical accuracy of between 0.01 and 0.05 mm in placement of sutures. In some cases, suture as fine as 13 μm has been used.

There have been various modifications of the suturing technique. Sutures are placed in either an interrupted or continuous manner. Interrupted sutures (Figure 3), usually 16 in number, tend to be slightly more irritating, stimulate more vascularization, and require earlier removal. However, this vascularization provides more rapid healing and earlier suture removal is possible and often safer. Continuous sutures are less irritating, but if the suture pulls through the tissue or loosens, the


1
2
3
4
6
5

entire wound is in jeopardy; it cannot be partially removed safely if one area of the wound has early vascularization. For this reason, continuous sutures tend to be reserved for those corneas with avascular opacities or thinning, and when interrupted sutures are used, cutting the suture on the knot and burying it minimizes irritation and gives results similar to continuous sutures.

The fine, less irritating suture has also been a double-edged sword, in that healing takes longer. Although its tensile strength gives a tight bond, healing may take up to a year, and removal of the suture even at that time has lead to wound dehiscence from lack of complete union between donor and host (8). This may cause graft opacification from either rejection or infection (9, 10).

Because slow wound healing requires sutures to remain in place over a long period, delaying visual rehabilitation, new suturing techniques have been devised. We are using a "double continuous" suture technique on avascular corneas. A 10-0 monofilament nylon suture is placed for watertight bond and a second, continuous 16-μm suture is placed less tightly. The first "coarse" suture is removed at about three months. Removing this tight suture releases tension, while the slightly less taut 16-μm suture keeps apposition to allow continued healing. Spectacles or contact lenses may be fitted for vision, as the fine nature of the 16-μm suture allows it to be covered with corneal epithelium and goes unnoticed by the patient.

Besides instrumentation, another critical factor in the improved prognosis for clear corneal transplants is a better understanding of corneal biology. With the awareness of the importance of the endothelial monolayer in maintaining corneal clarity and transparency, greater care is now taken to insure its integrity (11, 12). A damaged endothelium appears not to heal by cell division, but only by spreading of available cells, which are limited in amount. Endothelial damage allows the aqueous humor to penetrate into the corneal stroma, resulting in increased corneal thickness and graft opacification. The endothelium is extremely sensitive to the most minimal mechanical trauma, dessication, or chemical trauma such as oxygen deprivation or pH changes. It was common practice in the earlier days of transplantation to drive the needle into the donor cornea after wrapping the graft in gauze, to use clamps to hold it, or even place it on a towel endothelial-side down to get traction. All this was done using coarse suture and large, blunt needles, and could irremediably damage the delicate endothelial monolayer. This is in contrast to the use of fine instruments and particular attention now paid to minimizing endothelial trauma (13). It is for this same reason that donor corneal buttons are now commonly taken

Figure 3 Trephine is used to start the incision in the host cornea after the host eye has been immobilized by the prosthesis shown. The incision is deepened with a fine scalpel (2), and the cornea is lifted out while being fully excised with right-hand scissors along one side and left-hand scissors along the other (3). The eye is now ready for transplant (4). In aphakic eyes, vitreous humor must be cleared away, and this is done with the vitreous sucker-cutter (5). Suturing is begun with a fine curved needle (6). Normally (7), eight long single stitches are placed and then, after the cornea is sewn in with finer materials in zigzag fashion (color), the single stitches are removed. In the more difficult cases, 16 long single stitches provide a stronger seal (8). Reprinted with permission from *Hosp. Pract.* 8:69–75 (1973).

from corneal-scleral discs that have been carefully dissected from the whole eye. These corneal-scleral donors are placed endothelial-side up on Teflon blocks and cut with ultrasharp trephines. Careful study has shown that this technique causes less endothelial damage and gives more perpendicular edges than cutting from a whole eye, with either a complete trephine cut or partial trephine cut followed by scissors, from the epithelial side (14). In addition, the more perpendicular edges produced allow better wound apposition.

This greater respect for the corneal endothelium has revolutionized the grafting into eyes in which the lens (usually cataractous) has been removed. In the 1950s these aphakic eyes were considered almost hopeless for transplantation. However, the realization that damaged endothelium caused graft failure and the idea that the vitreous gel came forward in many of these eyes, damaging the endothelium led to removal of vitreous humor in aphakic eyes receiving transplants. Vitreous humor removal was to prevent subsequent vitreous contact with the endothelium and damage. Vitrectomies were initially performed by using syringes to withdraw "liquid vitreous" from the posterior portion of the eye (15). However, the vitreous also has a fibrillar, more solid gel that often protrudes into the anterior portion of the eye, producing endothelial contact. More recently, vitrectomies were performed on this gel, using cellulose sponges to which the gel-like vitreous humor adheres and then removing it by excision with scissors (16, 17). This exerts traction on the retina (18), is often tedious, and particles of the sponges flaking off into the eye can cause marked inflammation. New instrumentation has made vitrectomy easier. A suction-cutting device with a rotating blade to cut the vitreous fibriles can facilitate vitreous humor removal (19). It is used through the pupil after the cornea is removed, in either aphakic patients or patients undergoing simultaneous transplant and cataract extractions.

Two other instruments have been developed that improve the monitoring of the postoperative course of the graft. The first is an electronic applanation tonometer that allows accurate measurement of intraocular pressure in those patients with both irregular, scarred corneas (20) and those with keratoplasties (21). Increased intraocular pressure causes constriction of the vasculature around the optic-nerve head, leading to optic atrophy. This increased pressure postoperatively can also blow out the corneal wound. It is now realized that aphakic eyes and eyes undergoing combined cataract extraction and transplant are subject to elevated pressure. Prior to the advent of this instrument to measure the pressure in eyes with grafts, one had to measure it by touch or older tonometers, both of which are so inaccurate as to be grossly misleading. Now detection of increased pressure can prevent ocular damage with appropriate medical therapy to lower it. Pharmacologic control of this increased pressure can be achieved with drugs such as acetazolamide, which decreases aqueous humor production; parasympathomimetics such as pilocarpine, which increase outflow of aqueous humor; and hyperosmotic agents such as glycerol. In patients who do not respond to medical therapy or respond inadequately, external cryotherapy to the ciliary body (which produces aqueous humor) is a nonsurgical approach that lowers pressure without damaging the transplant (22).

This latter method of pressure-lowering is also used preoperatively to improve the prognosis for graft clarity in patients previously considered to be inoperable because of the inevitable failure in grafts in glaucomatous eyes.

The second instrument used postoperatively in the postoperative assessment of the transplant is the corneal pachometer (23). This is an optical device that attaches to the slit-lamp microscope that allows accurate measurement of corneal thickness to 0.02-mm precision. The transplant that is viable will be clear and thin postoperatively, or will exhibit a pattern of gradual thinning from a somewhat thickened state. Low-grade inflammation affecting the endothelium or other endothelial damage that allows subtle thickening of the transplant by aqueous-humor absorption may often be missed without such an instrument. Thickening of a thin transplant is often the earliest sign of endothelial decompensation or graft rejection.

Since there does not appear to be significant endothelial regeneration in humans, only sliding of cells to cover a deficit or cell damage, one can see why the protection of the endothelium in corneal transplantation is of paramount importance. It has been known for many years that the corneal epithelium rapidly regenerates, as with corneal abrasions. Because of this, little attention was paid to its handling during corneal transplantation. Epithelium was often removed from the donor with the idea of decreasing antigenicity or preventing the transmission of infection. Generally, the host epithelium covers the surface of a cornea within several days when the donor epithelium has been removed. It will also grow over fine suture. However, when the host has poor regenerative processes, the graft without epithelium can be compromised. This is especially true for patients with eyes scarred from a chemical injury. Damage to the cornea and conjunctiva, as well as to the lacrimal system, leaves the host with poor ability to epithelialize a graft. Chronically de-epithelialized corneas dehydrate due to evaporation, and further ulcerate. The abnormal epithelium produces collagenases that cause further ulceration of the graft (24). This problem with epithelialization is not only true of chemical-burn patients, but of patients with dry eyes (keratitis sicca) and poor lid function or scarring. Scarred, roughened lids promote abnormal drying patterns of the cornea due to improper surfacing of the tear film with blinking. In the past, patching or temporary suturing with the eyelids closed was the only treatment (if a transplant was even considered in these patients). However, the advent of the soft contact lens has revolutionized this aspect of corneal transplantation (25). The bandage–soft contact lens is placed in the eye postoperatively, and by its hydrophilic properties aids in moisturizing the dry eye and acts as a barrier to abrasion. The lens also relieves the discomfort of numerous sutures and smooths the corneal surface that may be abnormally dry due to tight wound closure.

Chemical-burn and dry-eye patients are difficult problems in corneal transplantation because the burned recipient epithelium will not heal properly and the donor epithelium must survive. Close medical attention and diligent care of soft contact lenses are needed to protect it. Yet it is only the recent realization of the importance of protecting the corneal epithelium and the promoting of its rapid healing with soft bandage lenses that has removed these types of patients from being hopelessly blind.

Besides the technical advances that have improved the prognosis in corneal transplantation, and the greater understanding of corneal immunology, the use of corticosteroids to attenuate and treat episodes of graft rejection has also been a significant factor in preventing graft failure. The cornea is a relatively immunologically privileged site because it lacks vascular and lymphatic supply (26). The normal avascular state of the cornea and the absence of lymphatics may protect the transplant from sensitizing the host as well as isolating the transplant from components of the immune system.

The new technology has decreased the trauma of surgery (Figure 4). The finer sutures have decreased the inflammation and postoperative vascularization, a prelude to graft rejection, of the transplant. The suppression of vascularization and inflammation is the ultimate goal in prevention and treatment of the graft rejection. To date, corticosteroids have proven to be the most effective drugs to do this (27). Topical corticosteroids are generally used to suppress inflammation and early vascularization, but systemic corticosteroids may be of value in conjunction with topical administration in patients with severly vascularized corneas to reduce initial sensitization as well as suppress inflammation. Immunosuppressive agents have been tried in corneal grafting, but their risks and toxicity appear to outweigh their benefits.

Evidence has accumulated that cell-mediated immunity is the predominant factor in the graft-rejection response. Humoral immunity and circulating antibodies may also be involved, and although potentially harmful, may also block sensitization and reactivity by the cellular immune system in the lymphatic-free cornea. It has been shown that all three corneal layers (epithelium, stroma, and endothelium) may react separately in the immune response (26).

Two specific areas of research are being investigated with regard to decreasing the incidence of graft rejection. Tissue typing and the matching of histocompatibility antigens are being pursued because of the great promise they have shown in renal transplantation (28). Because of the difficulties with matching tissue and because of the good prognosis of most corneal transplants, this technique may only prove useful for those severely vascularized corneas with poor prognosis. The ability to do this typing is made possible because of the availability of cryopreserved corneas for transplantation, which is discussed later.

The other research area being explored to prevent graft rejection is the use of blocking antibodies. It is postulated that graft antigen in the lymphatic-free cornea goes first to the spleen, making blocking antibodies that protect the normal graft and give it its immunologically privileged position. Exogenous blocking antibodies clearly delay and prevent rejection in animals, perhaps by binding to antigenic sites on the donor and preventing sensitization of the host (29, 30). Although this has not yet been tested in man, all it requires is dipping the donor button in purified HLA antibodies and virtually no risk is involved.

While advances in technology and corneal biology, and the use of corticosteroids have improved the prognosis of corneal transplantation, advances in the preservation of donor material have changed the "emergency surgery" aspect of keratoplasty. In the early years of transplantation, at the beginning of the twentieth century, fresh corneas from enucleated eyes were used. In 1934 Filatov was the first to describe the use of cadaver eyes (31). The cadaver eye stored short-term in refrigerated moist chambers has been the standard donor means of preservation. There is little hard data on acceptable time limits, from death to enucleation, for use of corneas. Generally accepted usage time for corneas is probably within 12–15 hr from cadavers that have been refrigerated immediately. The length of time the donor eyes are viable for penetrating keratoplasty when immediately enucleated and stored at 4°C (refrigeration) in a moist chamber is generally accepted as 24–48 hr (32).

Long-term storage of corneal tissue was made popular in the mid-1960s with the successful development of a reliable method for cryopreservation of corneas by Kaufman and associates (33). Careful control of factors to permit corneal endothelial survival made this technique successful (34). This ability to freeze tissue for transplantation would also allow time for the matching of histocompatibility antigens to be a viable technique for those especially poor-prognosis grafts. Tissue-culture techniques are also being studied, but contamination is common and the techniques complex. Cryopreserved tissue has proved viable for transplantation even one year after preservation (35).

Another, simpler method for corneal preservation that has been developed is intermediate between the use of fresh tissue (short-term storage) and cryopreserved tissue (long-term storage). An intermediate storage medium was developed by McCarey & Kaufman in 1974 (36). This preservation medium is made with 199 tissue-culture medium, 5% dextran, and antibiotics for prevention of infection. Corneas have successfully been grafted when stored up to six days in this medium (37). Even in short-term use, tissue stored in this nutritive solution, as opposed to being left on the whole eye, seems superior. After death the endothelium is bathed by aqueous humor whose nutrition is rapidly depleted and products of dying tissue are discharged into it. A nutritive solution obviates this.

Yet another innovation in the realm of donor material is in the evaluation of the donor eye. Besides the usual considerations of age of the donor, cause of death, and examination with slit-lamp microscopy for evidence of gross endothelial degeneration, the donor eye can now be examined by specular microscopy to see in situ the endothelial morphology and density of the corneal endothelial cells (38) (Figure 5). Whereas gross aging changes can be seen with the slit lamp, this instrument allows more subtle changes to be observed and screening donor eyes with it may yield a better quality of donor material. This same instrument, adapted for clinical use, also allows the observation of the corneal endothelium in vivo (39, 40). This may allow early detection of endothelial problems and possibly predict the ability of the cornea to withstand trauma or the inflammation of a rejection episode.

CONCLUSION

Advances in technology, methodology, corneal biology and immunology, and donor storage have radically improved the prognosis for corneal transplants as well as

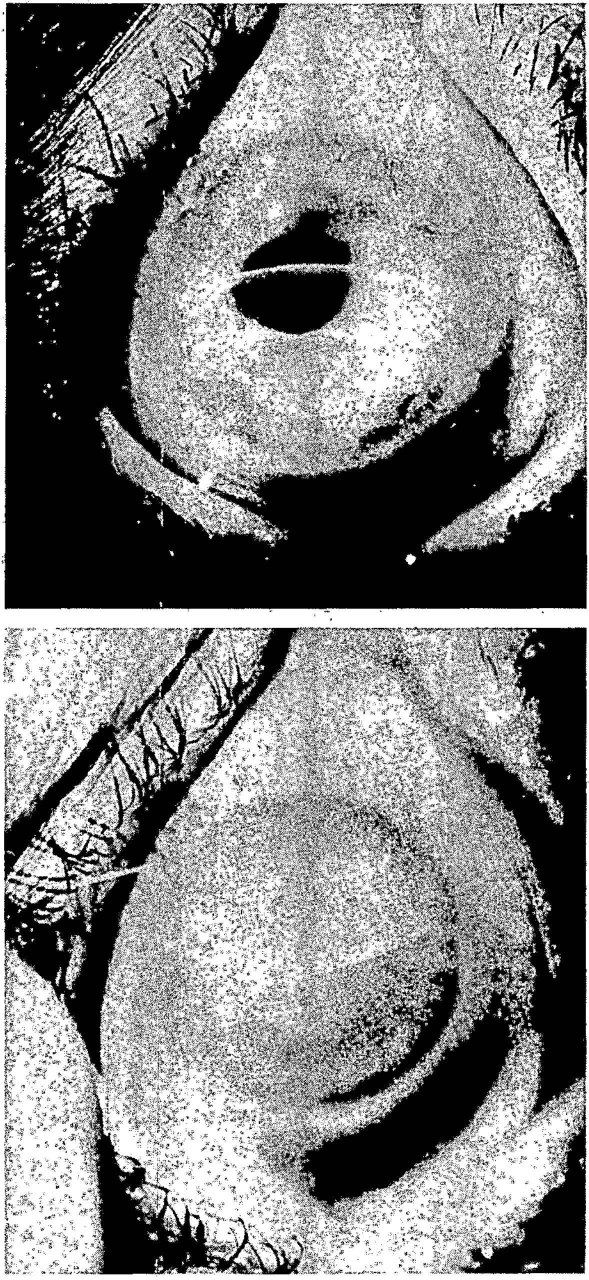


Figure 4 The corneal opacity (*left*) has been replaced by a perfectly functioning transplant (*right*) one year after surgery. Reprinted with permission from *Hosp. Pract.* 8:69–75 (1973).

changed the impact of surgery on the patient. In the past, patients would be "on call," waiting for a donor cornea often for weeks. Now they can be scheduled at optimal times for transplantation. The surgery can easily be performed under local anesthesia, and the elderly and patients with cardiovascular or respiratory problems can be considered for corneal transplantation. Age and debility will not deprive octagenerians and nonagenarians of a sight-restoring operation, since morbidity is insignificant and no limitation of activity is necessary. These new developments permit operation on children with congenital corneal opacities as early as several months in order to try to prevent the permanent visual disability due to visual deprivation.

The surgical procedure allows the patient to be up the same day for bathroom privileges. The eye can be unpatched the day after surgery. Within reasonable limits, the patient can return to normal activity and work, usually within a week. The time that sutures must remain varies, depending on the wound healing. This is usually quite early in children (about four weeks), while in the adult with an avascular cornea it may be six months to one year postoperatively. However, new suturing techniques use exquisitely fine suture (16-μm diameter) that may remain indefinitely and may allow excellent vision as early as 3 months, without fear of wound dehiscence due to an incompletely healed graft. Soft contact lenses have also helped in the visual rehabilitation through full-time or intermittent wear.

One can see from the table tabulating prognosis (Table 1) that the state of the art has come a long way in the 70 years since the first successful penetrating corneal transplant, and the "audacious fantasy" is truly reality.

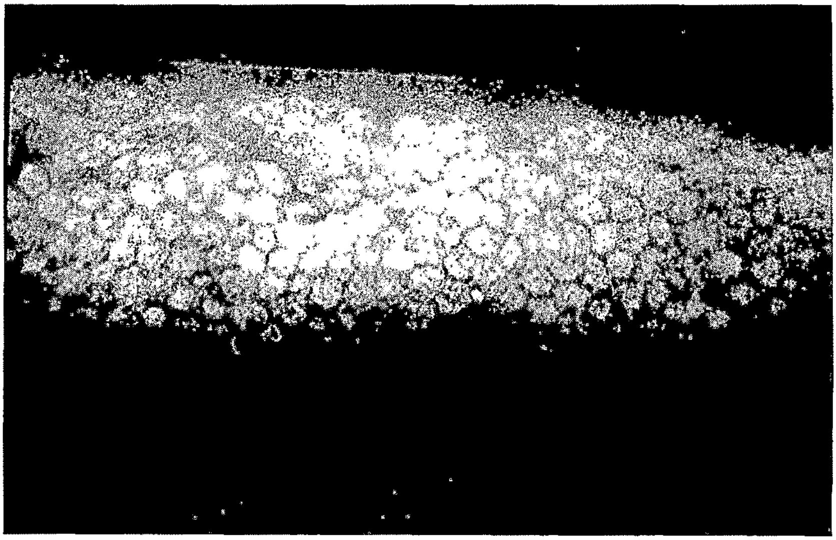


Figure 5 Specular micrograph of postoperative penetrating keratoplasty patient.

Table 1 Estimate of prognosis in penetrating keratoplasty

	Types of disease and prognosis	
Excellent (90+%)	Good (70-80%) (Phakic or Aphakic)	Fair (50-70%)
lattice, macular, and granular dystrophies keratoconus moderate corneal edema: Fuchs', etc	marked corneal edema: (primary or secondary severe dystrophy) inactive herpetic scar vascular scar vitreous touch controlled glaucoma or other complication	lye burn severe vascularization and/ or keratitis sicca, Stevens-Johnson active infection (surgery to salvage) poorly controlled glaucoma or major complications

Literature Cited

1. Zirm, E. 1906. Eine Erfolgreiche Total Keratoplastick. *Von Graefes Arch. Ophthalmol.* 64:580–93
2. Paton, R. T. 1955. *Keratoplasty,* pp. 4–35. New York: McGraw-Hill 280 pp.
3. Rycroft, B. W. 1955. *Corneal Grafts,* pp. 1–10. London: Butterworth. 285 pp.
4. Leigh, A. G. 1966. *Corneal Transplantation.,* pp. 1–5. Oxford Engl: Blackwell. 342 pp.
5. Trevor-Roper, P. D. 1972. In *Corneal Grafting,* ed. T. A. Casey, pp. 1–10. London: Butterworth. 340 pp.
6. Elschnig, A. 1930. Keratoplasty. *Arch. Ophthalmol.* 4:165–73
7. Filatov, V. P. 1935. Transplantation of the cornea. *Arch. Ophthalmol.* 13: 321–47
8. Brown, S. I., Tragakis, A. M. 1971. Wound dehiscence with keratopolasty: Complications of continuous-suture technique. *Am. J. Ophthalmol.* 72: 115–16
9. Binder, P. S., Abel, R., Polack, F. M., Kaufman, H. E. 1975. Keratoplasty wound separation. *Am J. Ophthalmol.* 80:109–15
10. Forstot, S. L., Abel, R., Binder, P. S., 1975. Bacterial endophthalmitis following suture removal after penetrating keratoplasty. *Am. J. Ophthalmol.* 80: 509–12
11. Harris, J. E., Nordquist, L. T. 1955. Hydration of cornea. I. The transport of water from the cornea. *Am. J. Ophthalmol.* 40:100–11
12. Dikstein, S., Maurice, D. M. 1972. Metabolic basis of fluid pump in the cornea. *J. Physiol.* 221:29–41
13. Jones, B. R., Rice, N. S. C. 1969. *In Corneo-Plastic Surgery,* ed. P. V. Rycroft, pp. 307–14. Oxford Engl: Pergamon. 628 pp.
14. Brightbill, F. S., Polack, F. M., Slappey, T. 1973. A comparison of two methods for cutting donor corneal buttons. *Am. J. Ophthalmol.* 75:500–6
15. Fine, M., 1960. Therapeutic keratoplasty. *Trans. Am. Acad. Ophthalmol. Otolaryngol.* 64:786–808
16. Kasner, D. 1968. Vitrectomy: Approach to management of vitreous loss. *Highlights Ophthalmol.* 11:304–29
17. Sanders, N. 1968. New treatment for bullous keratopathy: Penetrating graft and radical anterior vitrectomy. *Highlights Ophthalmol.* 11:296–303
18. Forstot, S. L., Binder, P. S., Fitzgerald, C., Kaufman, H.E. 1975. Incidence of retinal detachment after penetrating keratoplasty. *Am. J. Ophthalmol.* 80: 102–5
19. Brightbill, F. S., Kaufman, H. E., Levenson, J. E. 1973. Vitreous suction cutter for aphakic keratoplasty. *Am. J. Ophthalmol.* 76:331–35
20. Kaufman, H. E., Wind, C. A., Waltman, S. R. 1970. Validity of MacKay-Marg electronic applanation tonometer in patients with scarred irregular corneas. *Am. J. Ophthalmol.* 69:1003–7
21. Wind, C. A., Kaufman, H. E. 1971. Value of M-M applanation tonometer

following penetrating keratoplasty in man. *Am. J. Ophthalmol.* 72:117–18
22. Wood, T. O., West, C., Kaufman, H. E. 1972. Control of intraocular pressure in penetrating keratoplasty. *Am. J. Ophthalmol.* 74:724–28
23. Mishima, S., Hedbys, B. O. 1968. Measurement of corneal thickness with Haag-Streit pachometer. *Arch. Ophthalmol.* 80:710–13
24. Brown, S. I., Weller, C. A. 1970. Pathogenesis and treatment of collagenase-induced diseases of the cornea. *Trans. Am. Acad. Ophthalmol. Otolaryngol.* 74:375–83
25. Brown, S. I., Tragakis, M. P., Pearce, D. B. 1972. Corneal transplantation for severe alkalai burns. *Trans. Am. Acad. Ophthalmol. Otolaryngol.* 76:1266–74
26. Silverstein, A. M., Khodadoust, A. A. 1973. In *Corneal Graft Failure, Ciba Found. Symp.*, pp. 105–25. Amsterdam: Assoc. Sci. Publ. 363 pp.
27. Maumenee, A. M. 1973. See Ref. 24, pp. 241–55
28. Gibbs, D. C., Batchelor, J. R., Werb, A., Schlesinger, W., Casey, T. A. 1974. Influence of tissue-type compatability on the fate of full-thickness corneal grafts. *Trans. Ophthalmol. Soc. UK* 94:101–26
29. Chandler, J. W., Gebhardt, B. M., Kaufman, H. E. 1973. Immunologic protection of rabbit corneal allografts I. Preparation and in vitro testing of heterologous "blocking" antibody. *Invest. Ophthalmol.* 12:646–53
30. Chandler, J. W., Gebhardt, B. M., Sugar, J., Kaufman, H. E. 1973. Immunological protection of rabbit corneal allografts. *Transplantation* 17: 146–48
31. Filatov, V. P. 1934. Use of corneas from cadaver for transplantation material. *Sovet. Vestnik Oftal.* 4:221
32. McCarey, B. 1976. In *Current Concepts of Ophthalmology*, pp. 134–63. St. Louis: Mosby. 297 pp.
33. Kaufman, H. E., Capella, J. A., Robbins, J. E. 1965. Preservation of viable corneal tissue. *Arch. Ophthalmol.* 74:669–73
34. Kaufman, H. E., Escapini, H., Capella, J. A., Robbins, J. E., Kaplin, M. 1966. Living preserved corneal tissue for penetrating keratoplasty. *Arch. Ophthalmol.* 76:471–76
35. Kaufman, H. E., Capella, J. A. 1966. Preserved corneal tissue for transplantation. *J. Cryosurgery* 1:125–29
36. McCarey, B., Kaufman, H. E. 1974. Improved corneal storage. *Invest. Ophthalmol.* 13:165–73
37. Bigar, F., Kaufman, H. E., McCarey, B. E., Binder, P. S. 1975. Improved corneal storage for penetrating keratoplasties in man. *Am. J. Ophthalmol.* 79:115–20
38. Hoefle, F. B., Maurice, D. M., Sibley, R. C. 1970. Human corneal donor material. *Arch. Ophthalmol.* 84:741–44
39. Laing, R. A., Sandstrom, M. M., Leibowitz, H. M. 1975. In vivo photomicrography of corneal endothelium. *Arch. Ophthalmol.* 93:143–45
40. Bourne, W. M., Kaufman, H. E. 1976. Specular microscopy of human corneal endothelium in vivo. *Am J. Ophthalmol.* 81:319–23

Ann. Rev. Med. 1977. 28:37–42
Copyright © 1977 by Annual Reviews Inc. All rights reserved

IgA-ASSOCIATED GLOMERULONEPHRITIS

❖7220

John J. McPhaul, Jr., M.D.

Clinical Research Laboratory, Wilford Hall USAF Medical Center (AFSC), Lackland AFB, Texas 78236, and the Department of Internal Medicine, University of Texas, Southwestern Medical School and Veterans Administration Hospital, Dallas, Texas 75216

INTRODUCTION

Experimental immunopathologic studies in laboratory animals have defined the principal pathogenetic mechanisms causing glomerulonephritis: nephrotoxic (glomerular basement membrane) antibody and immune-complex diseases (1). Although these mechanisms have been confirmed in man (2), numerous clinicopathologic groups of human glomerulonephritis are incompletely defined and understood. Conspicuous among the latter is immunoglobulin A (IgA)-associated glomerulonephritis.

First described by Berger in 1969 (3), the salient immunohistochemical features constitute its most enigmatic facets: (*a*) diffuse and generalized involvement of every glomerulus with IgA deposits, usually with other immunoglobulins; (*b*) primary mesangial localization; and (*c*) indolent and persistent clinical disease, usually with hematuria, unassociated with other obvious systemic signs.

The apparent incidence of this glomerulonephropathy varies in different series from 5–20% of patients biopsied—the variation is probably related to different patient populations and indications for renal biopsy. Males are more frequently affected than females, and most patients are adolescents and young adults below 30 years of age. Although there is no unanimity regarding familial incidence of this disorder, there is unexpectedly frequent reference to clinical renal disease among family members in several reports (4–6).

IgA-ASSOCIATED GLOMERULONEPHRITIS

Clinical Syndrome: Clinical Features, Epidemiology, and Natural History

IgA-associated glomerulonephritis is diagnosed most frequently by renal biopsy undertaken to define the cause of macroscopic hematuria, with often recurrent and/or persistent microhematuria not explicable by other techniques. Although several reports suggest that this disorder is the most frequent cause of idiopathic hematuria, other reports find more varied immunopathologic bases for this syndrome (7, 8). It is important to stress that the diagnosis is established only by immunohistochemical staining of renal tissue positive for IgA.

At the time of diagnostic studies, most patients have normal renal function, trivial to modest proteinuria, and no signs of systemic disease, although 10–15% may have elevated serum creatinine concentrations and progress to renal failure. Nephrotic syndrome, arterial hypertension, and recurrent urinary-tract infections have been described in patients with this nephropathy, but are not usual. Several patients with IgA-associated nephropathy and renal failure have received renal allografts, and in numerous grafts recurrent IgA deposition has been observed (9).

Etiology and epidemiology of this clinical entity are obscure. Although an occasional patient has been documented to have intercurrent or recent streptococcal infection with first signs of disease, most give no discrete history of antecedent respiratory or skin infection. Nevertheless, gross hematuria, the single most frequent sign and symptom, often occurs within 24–72 hr of common respiratory infections, fever, unusual exercise, or trauma; an inciting event cannot be suggested in many cases. Although IgA-associated glomerulonephritis was reported first from France, subsequent studies have detected it worldwide.

Histopathology

The histopathology of renal biopsies in this disorder is varied. Although some biopsies are relatively normal or show only mild glomerular stalk thickening by light microscopy, most reports characterize the glomerular changes as a focal and segmental proliferative glomerulonephritis, primarily involving mesangial cells and matrix (10). In 20–30% of such cases, adhesions to Bowman's capsule, segmental glomerulosclerosis, and/or small fibrocellular crescents may be seen; focal necrosis of glomerular tufts is uncommon but has been reported. Diffuse and generalized mesangial proliferative glomerulonephritis has predominated in some series; tubulointerstitial changes that are noted are usually found in this group (3) and in patients with end-stage glomerulonephritis with extensive glomerular sclerosis. Patterns more typical of membranoproliferative and lupus nephritis are rare (11).

Electron Microscopy

Ultrastructural analyses by electron microscopy almost uniformly show mesangial deposits, commonly just below the glomerular basement membrane reflecting over the mesangial region, as well as mild mesangial hypercellularity and hypertrophy of mesangial matrix. In general, the number and size of mesangial electron-dense deposits correlate with the extent and intensity of mesangial staining of immunoglobulins seen by immunofluorescence. Although deposits along peripheral capillary loops are unusual, both subepithelial and subendothelial glomerular deposits have been reported; the dense deposits seen rarely in Bowman's capsule are of uncertain significance (12).

Immunofluorescent Findings

By definition, immunohistochemical stains show the characteristic mesangial staining of host IgA, usually accompanied by IgG and sometimes by other immunoglobulins; distribution of IgA is generalized and diffuse, and surprisingly intense even in biopsies that show no substantial abnormality by light microscopy. Stains for serum-complement components usually indicate deposition of C'3 and properdin, rarely C'_1q or C'4. Fibrinogen/fibrin is found often in a focal and segmental mesangial location, with much more variability between glomeruli than other proteins. Mounting evidence indicates that secretory IgA is not present in affected glomeruli.

FEATURES OF SPECIAL INTEREST

Interest in IgA-associated glomerulonephritis is directed toward three intriguing facets of its immunopathogenesis that have not been explained completely: (*a*) the role of IgA, (*b*) involvement primarily of renal mesangium, and (*c*) relationship of this syndrome to Henoch-Schönlein purpura (H.S.P.) and other systemic diseases with glomerular IgA deposition.

Role of IgA

IgA, a major component of the serum Ig pool, exists in two subclasses: IgA_1 is the major component of total serum IgA, whereas IgA_2 is in much greater concentration in external secretions (tears, nasal secretions, saliva, succus entericus, etc). IgA functions as the principal antibody class in these secretions, particularly against viral antigens, whereas the immune function of serum IgA is not as well specified. Although discrete antibody activities have been measured in the IgA class, including autoantibodies such as antinuclear and anti-gamma globulin factors, predominantly IgA systemic serologic response is not characteristic of usual immunogens (13). This disparity has prompted the suggestion that an antigen may diffuse from the gut and excite a preferential IgA response with resulting deposition of immune complexes in the kidney. In fact, several patients with IgA-associated glomerulonephritis have a history of chronic or recurrent sinus infections, and onset of hematuria in many patients seems triggered by common respiratory infections. Thus far, no evidence has been adduced in support of the alimentary antigen(s) or route of immunization, and the weight of present data does not favor significant participation of secretory IgA in this disorder.

Renal Mesangium

Important immunologic and ultrastructural studies have addressed the functional role of renal mesangium and factors contributory to the localization and deposition of circulating immune complexes; they have been reviewed recently (14, 15). Mesangial function in transport and disposition of a variety of macromolecules, including colloids and protein aggregates, has been shown to be a dynamic process, with apparent preferential collection along subendothelial aspects of glomerular basement membrane, uptake into the mesangium, and movement in channels between mesangial cells towards the hilum and vascular pole of the glomerulus. The loss of electron-dense, fluorescent, or radioactive markers in these experiments is interpreted to result from degradation of the test materials and transport into hilar lymphatics.

Immune complexes formed in vitro and administered passively, as well as complexes formed in vivo in response to soluble protein and viral antigens, may localize preferentially to the mesangium. Such mesangial deposition is characteristic of larger complexes with very high molecular weight antigens or complexes formed in relative antibody excess. Corticosteroids can influence this localization by altering antibody synthesis and affecting antigen:antibody ratios, by affecting metabolism of complexes, or by direct effects on GBM and/or the mesangium (16).

Although the renal mesangium functions in the handling and ultimate fate of a variety of macromolecules that may not be immunologically active, it clearly is an important site—even a primary site—of immune-complex localization.

Relationship to Systemic Diseases with IgA Deposits

Primary mesangial deposition of serum immunoglobulins, including IgA, has been noted in systemic diseases, particularly in H.S.P., systemic lupus erythematosus, viral hepatitis, and hepatic cirrhosis. In many cases the glomerular immunoglobulin is predominantly IgA, such as in H.S.P. and hepatic cirrhosis; the putative antigen(s) involved that elicits presumed IgA antibodies is not evident.

Significant elevations of serum IgA concentrations have been noted in hepatic cirrhosis and H.S.P., particularly early in the disease (17). The frequent clinical involvement of gastrointestinal tract in H.S.P., increased serum IgA concentration, and renal deposition of IgA has focused attention on the GI tract as source of antigen and/or antibodies in this presumed immunological disease. While serum IgA concentrations have been reported to be high in IgA-associated glomerulonephritis without systemic disease, not all series are in agreement.

Further data linking IgA nephropathies with or without systemic disease have been offered in reports that IgA and other immunoreactive proteins can be seen in walls of cutaneous capillaries of uninvolved skin of patients with IgA-associated glomerulonephritis and patients with H.S.P. (18, 19). Confirmation of these observations would add to the likelihood that IgA deposits in both sites resulted from circulating immune complexes, although systematic use of tests for detecting such complexes in IgA-associated glomerulonephritis have not been rewarding thus far. Moreover, the similarities of IgA-associated nephropathy to the nephritis of H.S.P. are so striking that it is not clear whether there is only a quantitative difference between the two disorders, or more fundamental differences.

IMMUNOPATHOGENETIC HYPOTHESIS

The primary hypothesis regarding immunopathophysiology operative in IgA-associated glomerulonephritis suggests that the nephropathy results from renal deposition of circulating immune complexes containing IgA. The preferential renal mesangial site of deposition is thought to result from the size of the circulating immune complexes; the indolent clinical pace of the nephritic lesion and its frequently unimpressive morphologic response probably result from a slower rate of immune-complex deposition and the faster turnover and disposal of complexes localized to the mesangium. In the absence of systematic data, it has not been demonstrated conclusively whether IgA is present as antigen or antibody; one study has indicated that the renal IgA does not have antiglobulin activity and that sera from these patients do not have anti-IgA antibodies (20). Cryoglobulins have been found infrequently in IgA-associated glomerulonephritis and have not always contained IgA. Although it is not likely in the usual case, anti-mesangial antibody participation has not been ruled out (20).

Aggregates of IgA are known to initiate complement activation through the "alternate" pathway (21). Numerous studies of renal biopsies from patients with IgA-associated glomerulonephritis attest to the high frequency of C'3 (Beta 1C) and properdin deposition in these glomeruli, although C'_1q and C'4 have rarely been seen; serum concentrations of classical and alternate pathway components as a rule are normal in these patients. The data are compatible with activation of the complement seen in these lesions through the alternate pathway. The majority of tests performed for cold activation of complement (22) and C'3 "nephritic factor" have been negative (23).

SUMMARY

In summary, IgA-associated glomerulonephritis is an interesting clinical problem. The immunohistochemical identification of renal IgA deposits is the sine qua non of its diagnosis, although most of the patients reported have had hematuric syndromes, particularly recurrent gross hematuria. The importance of this immunopathologic entity devolves from the crucial use of special stains to identify IgA, the enigmatic role of IgA, the usual mesangiopathic expression of histologic response, and the ill-defined relationship of this clinical problem to nephropathies associated with systemic diseases that also have glomerular IgA deposits.

Although still unproven, it is likely that the usual instance of IgA-associated glomerulonephritis is due to deposition of circulating immune complexes containing IgA. The nature of the exciting antigen(s), quantitative measures and characteristics of such complexes, and the role of mediating systems, including coagulation, have not yet been elucidated.

Literature Cited

1. Unanue, E. R., Dixon, F. J. 1967. Experimental glomerulonephritis: immunological events and pathogenetic mechanisms. *Adv. Immunol.* 6:1–90
2. Dixon, F. J. 1968. The pathogenesis of glomerulonephritis. *Am. J. Med.* 44: 493–98
3. Berger, J. 1969. IgA glomerular deposits in renal disease. *Transplant. Proc.* 1:939–44
4. deWerra, P., Morel-Maroger, L., Leroux-Robert, C., Richet, G. 1973 Glomérulites à dèpôts d'IgA diffus dans le mésanguim. *Schweiz Med. Wochenschr.* 103:797–803
5. Sessa, A., Cioffi, A., Allaria, P., Conte, F., D'Amico, G. 1973. Hereditary nephritis with immunofluorescent mesangial deposits. *Lancet* 2:853
6. Sissons, J. G. P., Woodrow, D. F., Curtis, J. R., Evans, D. J., Gower, P. E., Sloper, J. C., Peters, D. K. 1975. Isolated glomerulonephritis with mesangial IgA deposits. *Br. Med. J.* 3: 605–6
7. van de Putte, L. B. A., de la Riviere, G. B., van Breda Vriesman, P. J. C. 1974. Recurrent or persistent hematuria: sign of mesangial immune complex deposition. *N. Engl. J. Med.* 290:1165–70
8. Kupor, L. R., Mullins, J. D., McPhaul, J. J. Jr. 1975. Immunopathologic findings in idiopathic renal hematuria. *Arch. Intern. Med.* 135:1204–11
9. Berger, J., Yanevà, H., Nabarra, B., Barbanel, C. 1975. Recurrence of mesangial deposition of IgA after renal transplantation. *Kidney Int.* 7:232–41
10. Levy, M. Beaufils, H., Gubler, M. C., Habib, R. 1972. Idiopathic recurrent macroscopic hematuria and mesangial IgA-IgG deposits in children (Berger's Disease). *Clin. Nephrol.* 1:63–69
11. Finlayson, G., Alexander, R., Juncos, L., Schlein, E., Teague, P., Waldman, R., Cade, R. 1975. Immunoglobulin A glomerulonephritis: a clinicopathologic study. *Lab. Invest.* 32:140–48
12. McCoy, R. C., Abramowsky, C. R., Tisher, C. C. 1974. IgA nephropathy. *Am. J. Pathol.* 76:123–44
13. Vodopick, H., Chaskes, S. J., Solomon, A., Stewart, J. A. 1974. Transient monoclonal gammopathy associated with cytomegalovirus infection. *Blood* 44:189–95
14. Vernier, R. L., Mauer, S. M., Fish, A. J., Michael, A. F. 1971. The mesangial cell in glomerulonephritis. *Adv. Nephrol.* 1:31–46
15. Germuth, F. G. Jr., Rodriguez, E. 1975. Focal mesangiopathic glomerulonephritis: prevalence and pathogenesis. *Kidney Int.* 7:216–23
16. Haakenstad, A. O., Case, J. B., Mannik, M. 1975. Effect of cortisone on the disappearance kinetics and tissue localization of soluble immune complexes. *J. Immunol.* 114:1153–60
17. Trygstad. C. W., Stiehm, E. R. 1971. Elevated serum IgA globulin in anaphylactoid purpura. *Pediatrics* 47: 1023–28
18. de la Faille-Kuyper, E. H. B., Kater, L., Kuijten, R. H., Kooiker, C. J., Wagenaar, S. S., van der Zouven, P., Dorhout Mees, E. J. 1976. Occurrence of vascular IgA deposits in clinically normal skin of patients with renal disease. *Kidney Int.* 9:424–29
19. Tsai, C. C., Giangiacomo, J., Zuckner, J. 1975. Dermal IgA deposits in Henoch-Schönlein purpura and Berger's nephritis. *Lancet* 1:342–43
20. Lowance, D. C., Mullins, J. D., McPhaul, J. J. Jr. 1973. Immunoglobulin A (IgA) associated glomerulonephritis. *Kidney Int.* 3:167–76
21. Götze, O., Müller-Eberhard, H. J. 1971. The C_3-activator system: An alternate pathway of complement activation. *J. Exp. Med.* 134:90s–108s(Suppl.)
22. Day, N. K., Geiger, H., McLean, R., Resnick, J., Michael, A., Good, R. A. 1973. The association of respiratory infection, recurrent hematuria, and focal glomerulonephritis with activation of the complement system in the cold. *J. Clin. Invest.* 52:1698–1706
23. Zimmerman, S. W., Burkholder, P. M. 1975. Immunoglobulin A nephropathy. *Arch. Intern. Med.* 135:1217–23

Ann. Rev. Med. 1977. 28:43–55
Copyright © 1977 by Annual Reviews Inc. All rights reserved

CONFIDENTIALITY IN MEDICAL PRACTICE

❖7221

Maurice Grossman, M.D.
Department of Psychiatry, Stanford University School of Medicine,
Stanford, California 94305

For his patients' sake and increasingly for his own as well, the physician would be well advised to learn the narrow but tortuous path between the edict "reveal not your patients' data" and growing demands that he do just that.

BACKGROUND

The Greeks started it! Many medical concepts of the fourth century BC and after were attributed to Hippocrates. Some were products of his followers, with credit given to the Hippocratic school. The Hippocratic oath was essentially a code of ethics credited to the Pythagoreans. They promulgated a strict code of social and moral ethics in reaction to the less concerned standards of their society and of other medical groups (1). The more disciplined attitude, inferring that less rigid practices were current, suggests that during the fourth century BC physicians were not above gossiping and revealing their knowledge about patients to recipients whose concerns were not for the patient. The Hippocratic oath also adjured the physicians against giving poisonous drugs to those requesting such drugs. While Edelstein deduces that the intent was to forestall suicide (1), the language also suggests that the recipients of the poison might intend to administer it to someone else. The concept of "poison pen" material is not too farfetched in application to confidentiality.

In the closing quarter of the twentieth century AD, concern about physicians gossiping and releasing medical information to others is still with us. Currently, medical care involves many adjunctive professionals in supplying health services. They need access to this medical information. Social attitudes have changed—mostly superficially. Two major factors have not changed. Firstly, personal information is still necessary for determining health care in treatment and prevention of illness and trauma, and for research to improve the whole process. Secondly, some of the personal data can be poisonous to the patient, depending on who uses it and how it is used.

Everything would be simple, were we guided merely by the tenet "Do not reveal anything to anyone." Unfortunately, confidentiality cannot be absolute. The patient might require and therefore request revelation of some facts in order to get benefits based on the information. Others—family, schools, and even employers—could use some information to help the patient to overcome his disability. Here too, the patient should consent to the disclosures. There is divided opinion about whether it should be done without the patient's consent when he is not looking after his own health interests. There are situations where society's overriding needs and rights demand violating the patient's rights to privacy and, therefore, confidentiality (2). One must ask who determines these overriding rights of society, and what the basis is for that determination. These complications set the stage between the right to privacy and the need to know, with no clear delineation. As the tortuous path is followed, note that "the need to know" becomes "the right to know."

The Hippocratic concept is commingled with the popular attitude of society, accepted by many physicians, that medical information confided by the patient or discovered about the patient will never be divulged. This sacrosanct concept is supported by laws requiring such observance under threat of criminal and civil consequences for violation. As usual, this legal demand allows exceptions when the law wants such information (3). In 1828 New York was first to enact protection against legal demands for this sensitive data. New York was soon followed by most other states. The legal onslaught continued, with breaches in this protective barrier through nibbling exceptions created either by legislative act or judicial decision. As late as 1960, the major admitted threat to protecting patients from disclosure was the power of courts, legislatures, and other branches of government (4). The process of legal erosion has reduced this protection—physician-patient privilege—to virtual impotence (5).

BASIC DEFINITIONS AND CONCEPTS

Certain definitive differences must be considered in threading one's way through the problems. "Privacy" alludes to personal investments that must not be intruded upon by others. The individual may relinquish this right in many ways. In confiding some private aspects to another, the individual may explicitly or implicitly exact an obligation that the recipient keep this confidential. Thus "confidentiality" implies this obligation. As in medicine, it may imply that the disclosure was made to the second party to fill an urgent need. The second party has no proprietary right to the private material. If certain legal requirements are met, this confidential exchange can be protected legislatively from legal demands (6). This information is then privileged and leads to "physician-patient privilege," "psychotherapist-patient privilege," "attorney-client privilege," etc.

The above sequence from privacy to confidentiality to privilege does not alter the basic premise that the information is about an individual who alone maintains the proprietary right to that information. Having that right also leads to the corollary that that individual alone has the right to disclose this private information, while others do not. He may authorize other holders of the information to reveal it to third parties for whatever reason that impels the owner of the private material to issue the authorization.

A trap for the unsuspecting one is the surrender of privacy implicit in his making his data public. Usually the data subject authorizes release of information to secure certain benefits. This may be done willingly, e.g. to seek exemption from military service, or unwillingly, e.g. to obtain insurance benefits that will be withheld if permission for getting medical information is refused. The wording of authorizations may be so broad that the patient has waived all rights to forbid dissemination of the material. When the patient authorizes a release of information to a government agency, he may be unwittingly releasing it to files that are legally public records, available to anyone. In court, if the patient or his attorney makes reference to medical-record details, he may have waived privacy voluntarily, and thereby his right to privilege and its protection (7).

INSURANCE INDUSTRY INTRUSIONS

Legal demands for medical information, usually by subpoena (sometimes by search warrant), are relatively infrequent. However, since 1960 the growing increase in insurance coverage for medical-care costs has been accompanied by increasingly frequent requests for medical information to justify claims for payment (8). An insurance policy is a contract with specific limits and restrictions. The insurance carriers need information to determine that the claim is for services within the limits of the contract. Each claim form has a paragraph authorizing release of medical information, to be signed by the claimant. With few exceptions, the authorization is relatively unlimited in the breadth of disclosure the carrier can seek. It does not limit the dissemination of the disclosed information. It sets no criteria for the ultimate fate of the revealed information. Refusal to give this unlimited demanded information has resulted in refusal to pay claims that are otherwise justified.[1] The insurance carrier takes the position that it alone can determine how much and which information is necessary to justify claims payment. Medical information irrelevant to the cause for the claim has been used to justify the invalidation of the obligation, suggesting one reason for wanting an unlimited authorization (9). Some insurance companies do this with medical information obtained illegally (10).

Since most medical-care insurance is under group industrial policies, medical information released by patients to an insurance company has reached employers and, through employment channels, fellow employees. The risk to employment, promotion, character attacks, and emotional shock related merely to the public

[1] When Stanford University Hospital refused to permit a Blue Cross employee to be stationed at Stanford and given access to patients' records to determine appropriateness of length of stay, Blue Cross started delaying reimbursement. This forced acceptance of Blue Cross's demand in June, 1976. When Blue Cross of Georgia withheld payments when the College Street Hospital of Macon, Georgia refused their demand for complete access, the latter sued Blue Cross. Z. Sweeney Sikes, M.D., who pursued this, reports an agreed stipulation and verdict in February, 1975, awarding the hospital the approximately $70,000 in question.

exposure is greatest where psychiatric illness is involved, but not limited to that specialty of medical practice (11). The unlimited authorization permits such practice. In employer-administered group policies, the employee or his dependent patient may be obligated to authorize broad disclosure to the employer with no restriction of usage.

With the Federal Fair Credit Reporting Act requiring disclosure of files and sources of information, insurance companies have started informing applicants for policies that they use outside investigators—armed with a photostat of the authorization signed by the applicant—to get full medical records from all prior treatment sources, including unconfirmed information from neighbors, friends, family, employment sources, and wherever the claimant has contact. This information is stored without verification. The competition between these outside investigators leads to advertising the high percentage of adverse information they collect about investigated subjects. Since much of the information is unverified, it suggests what sort of bias may enter the collection. At least one company advertises that it stores this microfilmed data for future use without specifying to whom and on what grounds it will release the data (12). It might include records from hospitals and physicians, obtained supposedly with consent originally given to an insurance company.

However, some insurance companies will accept a simple general statement for a health-insurance claim; a verbal report to their medical director, avoiding records of sensitive data; or, in one plan, the mere statement "confidential illness" being checked obviates any other detail. When a large sum is involved, some companies will accept a peer-review decision, also avoiding written records. In July, 1976, the American Psychiatric Association and the Health Insurance Association of America began testing a code-reporting system formulated by Richard G. Johnson to avoid reporting identifiable material (13). The foregoing indicates that the problem is perpetuated by some companies and is not industry-wide. It also shows that insurance reporting can get along with less data than is prevalently "needed."

LEGAL PROFESSION'S DEMAND FOR DISCLOSURE

The demands of the law for disclosure of this confidential data has been mentioned in the preceding discussion of privilege on page 45: The original physician-patient privilege, more noted for its holes than for its protective fabric, never was acceptable to the legal profession. The severe destructiveness caused by public revelation in court of details gathered in psychiatric treatment stemmed from social attitudes about psychiatric illness as a category, the intimacy of the private details revealed during treatment, and the fact that such patients were already under intolerable emotional stress that led them into treatment initially (11). The tendency to view psychiatry as detached from medicine obscures the fact that these same problems of illness exist in every physician's practice, regardless of specialty. Every physician's patients may be subject to the same emotional disclosures and subsequent trauma. The only need for physician-patient privilege is for data that would affect the patient emotionally if revealed to others and would in effect create a psychiatric syndrome. Courts have recognized that the protection of psychotherapeutic communications is based on constitutional rights, as well as meeting the criteria for privilege (14). Unfortunately, such legislation is still limited to California, Connecticut, Florida, Georgia, Illinois, Kentucky, and Massachusetts. Worse, they limit the protection to patients of psychiatrists and psychologists. Rule 504 of the proposed Federal Code of Evidence (5) would have made the protection available to all physicians' patients. The legal profession's growing boundaries of the right of discovery, ostensibly to bring in all the facts available to reach a just verdict, even denies this protection in the exception when the patient uses his emotional condition for a claim or a defense. It makes no leeway for circumstances where the needed facts can be obtained other than through the treatment source, even if these other sources may also be physicians. Parenthetically, the legal profession's virtuous search for the truth, breaching constitutional rights of the patient, does not extend to their own files and the truths confided to them by their clients. The legal profession's use of subpoenas and how physicians are intimidated by them to reveal information unnecessarily has been discussed elsewhere (15).

This exception to privilege compounds the insurance problems of patients and physicians. Patients who exercise their rights to withhold unnecessary data and thereby lose their benefits, have recourse only to court action. In filing a claim, they lose the protection of privilege in that action because of the exception. Therefore, they lay themselves and their medical records open to the public—a "Catch-22" situation. Similarly, when physicians are slandered for protecting their patients' interests, they cannot go to court because it, too, would mean discussing their patients openly to explain that protection.

ALL MEDICAL COMMUNICATIONS ARE NOT CONFIDENTIAL

When it was pointed out in the second paragraph above that physicians other than the treating one could give the facts necessary, it demonstrated that all professional communications between an individual and a physician are not confidential. This term applies only to those communications that are given in confidence as a necessary part for treatment. Whenever a patient knowingly goes to a physician for a report to be made to some other party, there is no "confidence." In some cases, as in court-ordered examinations, the participation may not be voluntary. Examples are medical reports ordered by current or prospective employers, applicants seeking licensure that requires a medical examination for meeting standards, and medical recommendations required for admission to certain schools. Such examinations may be as minimal as a vision test for a driver's license or as extensive as the inclusion of a psychiatric examination where indicated. It has been recommended that the report be minimally sufficient to meet the objectives of the examinations (16). Such patients should be warned that any material divulged might become part of the report. Especially when a report might lead to an adverse finding, it is prudent to include details supporting that judgment. There is no agreement by the examiner for confidentiality. If anything, there is prior understanding that there is no confidentiality. Any agreement of confidentiality can be only between examinees and those requesting the examination. The availability of such examinations by others emphasizes that treating physicians should try to avoid supplying sensitive medical data for such purposes even when patient requests it. Therapists who do comply may jeopardize the protective climate for "privilege." They may alienate their patients if the medical report leads to an adverse finding. If they do so without patient's consent, except under legal duress, they may have placed themselves in jeopardy.

FAMILY MEMBERS' WISH TO KNOW

The most frequent requests for confidential information come from families of patients. Their interest in, concern for, and wish to help the patient are natural and to be expected. In most situations, sharing this information with the family is helpful and acceptable to the patient. There might be some information that the patient does not want to divulge even to the most dedicated member of the family. Physicians have learned that in some families the dedication to each other leaves much to be desired. Professed and superficial dedication may be belied by family behavior. Sensitive information divulged to a parent, child, or spouse has been used destructively against the patients. This requires a clear understanding that a patient truly consents to release information to a family member or specifies what information is to be withheld. Where there is family emotional involvement in a situation requiring treatment of one or more members, whether by a psychiatrist or other physician, it might be well to agree upon a policy about confidentiality at the onset —absolute, conditional, or none. The judgment should always be based on the general principle "Is breaking confidentiality, even with the family, for the patient's interest or destructive to the patient's interest?" (17).

COMPLICATIONS FROM CIVIL RIGHTS OF MINORS

The family relationship in confidential medical-data disclosure has always been complicated when the medical information is about minor children. The right of parents to this information has been eroded by developments of relatively recent origin. The erosion is partly due to growing awareness that children also have civil rights independent of parents. This is crucial when children avoid seeking needed medical help because they fear parental discovery and its consequence (18). Children have been known to commit suicide when faced with the dilemma. Recent legislative trends have permitted physicians to treat minors without parental consent. This has been legalized in forty states (18) and supported by law-enforcement action in specific instances. Removing the need for parental consent implies the child's right to privacy in the matter and the independent right to confidentiality. The need has been legislatively directed at treatment for venereal disease, pregnancy, contraceptive devices, alcohol, and other drug abuses. Those treating children seek wider application of the civil rights of minors. The physician will see the children for other problems, where violating the child's confidence would disrupt needed treatment or create emotional upheaval in some child or the family.

There is one rule to follow if it becomes important to accept the conditions of confidentiality and the child is so promised. The physician must be sure, and probably should so note, that the child is mature enough to understand the problem faced and to cooperate with treatment, and understands the full meaning and consequences of withholding information from parents. In counseling about the problem, a child might be helped to see that parental knowledge would advance the child's interests and not result in the feared adverse consequences. In borderline cases, with the child's consent, professional consultation might be helpful. Children usually go to a physician already familiar to them or trusted by them. A physician treating a family with children approaching puberty could seek agreement with parents and child that the physician have the consent of the parents to enter such a confidential relationship with the child in the future, if the child so decides at that later time. If the child is present when the agreement is reached, and knows this is available, it can prove to be good preventive medicine.

This extension of confidentiality illustrates another problem. In the first place, such treatment might require public health notification. It may be difficult to explain to the patient, adult or minor, that it is a legal requirement. Guaranteeing that a third party maintain confidentiality has proven to be risky. In the case of a minor, investigation of a communicable disease, and reports of pregnancy and abortion might lead to communication at the home address that will disclose the problem to parents. In making arrangements, this should be borne in mind. The question of preventing extension of the communicable disease to family members is also a problem.

DILEMMAS CREATED BY *TARASOFF* CASE OBLIGATION

Prosenjit Poddar, a student at UC Berkeley, was in psychotherapy on an outpatient basis at Cowell Memorial Hospital of the University of California in 1969. When Poddar began making threats to kill a fellow student, Tatiana Tarasoff, and Poddar's therapist, Dr. Lawrence Moore, believed that the threats constituted a possible danger, the latter asked the campus police to take his patient into custody incident to pursuing the legal machinery for involuntary hospitalization. This request was made on August 20, 1969. At the time Tatiana was in Brazil for a summer-long visit. The police did take him into custody, but released him when they found no sign of abnormality or of dangerousness.

Poddar caused no problems, even though he terminated treatment after his therapist took this step. He even became a roommate of Tatiana's brother. Shortly after Tatiana returned, Poddar did kill her, on October 27, 1969. Tatiana's parents filed suit against all individuals involved, and the Regents of the University of California as their employer, for a wrongful death. The plaintiffs gave four legal grounds in support of their claim. The second was based on the defendant's failure to warn the parents of the danger to Tatiana.

The lower courts dismissed the action without a trial because of legal immunity of the defendants plus the overriding need to maintain confidentiality in psychotherapy. The dismissal was appealed to the California Supreme Court, who reversed the lower court's decision and ordered the case to trial in a decision filed December 23, 1974. The Court stated that psychotherapists have a duty to warn potential victims when they determine that their patient is dangerous. Professional societies joined the defendants in a petition for a rehearing that was granted; the hearing was held May 5, 1975. The basis for the amicus curiae intervention of the professionals was that violating confidentiality under those circumstances would be a major impediment to effective treatment for all patients; and that in approximately 99 times out of a hundred the judgement of dangerousness would be invalid as proven by repeated studies. The Court filed its revised decision July 1, 1976, which did not alter the first decision in its essential feature obligating the warning.

Unfortunately, legal redress for untoward consequences of maintaining confidentiality are usually judged after the fact. This is in contrast to making a predictive judgment about breaching confidentiality, with its possibilities or probabilities, as the dilemma first presents itself. The ramifications of a California Supreme Court decision of potential hazard to a third party goes far beyond psychiatric treatment problems as indicated by the amicus curiae briefs in the petitions for, and subsequent rehearing in the *Tarasoff* case (19). The final decision destroyed an important element in protecting confidentiality. It changed the medical ethics (2) and legislative intent (20) from "no privilege" or "may disclose" to "must disclose," when the physician foresees a potential hazard to a third party. "The duty to warn" in that decision (19) ignores the difference between foreseeing a potential hazard and foreseeing an actual culminated destructive act.[2] In the field of psychotherapy the decision also fails to differentiate between the predictability of a psychogenic source of hazard that depends on a future unforeseeable external circumstance for triggering actual damage and a biologic source of hazard that is present in an infected patient and only requires physical transmission to create the actual damage. The decision also overlooks the realities of prevention in the two. Relevant to the duty to warn the potential victim encompassed by the decision, it can be demonstrated that the warning offers no protection and the court itself documents that it may increase the hazard. In cases of venereal infection in an adult or a minor, avoidance of contact is all that is necessary until current treatment rapidly renders the disease noncommunicable. In the view of the court, withholding of information from the parents can in some cases, after the fact, be alleged to have created emotional distress that a jury might later consider actionable damage. In terms analogous to the *Tarasoff* decision, should the physician refuse the child's request for confidentiality and disclose the information to parents, the possible damage of the child's suicide must be considered. An after-the-fact evaluation might again leave the physician in an actionable position.

---

[2] It is good medical practice to break confidentiality if a hazard to society demands it. The decision overlooks the fact that the greatest protection is the indicated treatment. Where this is failing, the California legislature determined, after extensive investigation, that due process for the therapist is as spelled out in the Lanterman-Petris-Short Act (21).

COMPUTERS COMPOUND THE PROBLEM

All the threats or fears of mishandled information have been compounded by the growing use of computers for depositing, storing, and retrieving information. The villain is not computer science, because its developments have contributed greatly to all sectors where information is important. This is true for medical science in all its branches, including treatment. The problem is too-ready access to the information with inadequate controls and anonymity for those tapping into the computer for confidential data. The ease of storing, transferring, and retrieving information has proliferated a demand for access to all information, including confidential medical information, for various purposes ranging from medical research to private industry wishes to governmental programs. Some of these demands have little to do with the medical welfare of the patients or society. In some cases the information is gathered secretly and used against the interests of the patient (22). They document the tie-in of such data from site of origin to local, county, state, regional, and national data banks.

That these concerns are not limited to psychiatric practice, or even just to the United States, is attested by the World Medical Association. Their annual two-day scientific program in Munich, Germany in October, 1973 was devoted exclusively to this problem. Speakers from all over the world discussed various facets of the problem. At least two nations with reputations of liberal orientation openly use identifiable medical data in their computer banks for all kinds of governmental purposes. The only delegates who thought this proper came from iron-curtain countries, and from public health officials of newly evolving countries trying to establish improved mass-delivered health services.

CONFIDENTIALITY AND MEDICAL RESEARCH

A recurrent theme during that conference is constantly at the surface of controversy in the United States. A large segment of researchers indicate that any restriction of access to data-subject identifiable medical information will interfere with needed research. They support, and some even demand, that all data generated at treatment sites be gathered in regional and national data banks (22d). This is the position taken by a research organization headed allegedly by John Baldwin in Oxford, England, operating the Oxford University Record Linkage System (according to several medical sources interviewed in London in October, 1975). The Task Force at the 1976 Key Biscayne Conference on Confidentiality of Health Records independently recommended controlled procedures (23), followed by the Medical Research Council, based at the University College Hospital Medical School, London, England (visited October, 1975). A variation used at St. Patrick's Hospital in Dublin, Ireland by R. J. Draper demonstrated the feasibility of using independent code numbers kept at the treatment site, with follow-up via treatment site, where identifying data is kept separately. The researcher is thus blocked from having or needing the identifying data. Those who devise such protective features in their protocol have uniformly expressed doubt that random collection of infinite details will be either manageable or rewarding, even with computer sorting or correlating. Yet the argument goes on internationally between those who hope that the computer will be the wished-for breakthrough if they feed it enough details and those who base their research on a planned protocol that may indicate some need for data-subject identifiable data.

PATIENTS' ACCESS TO THEIR RECORDS

A new area of complexity is the growing belief that patients—and parents of minors—have a right to all medical records to insure accuracy of historical data, as a check on treatment adequacy and to permit better cooperation in treatment (24, 25). This is already federal law, since the Privacy Act of 1974, P.L. 93–579, does not differentiate between medical data and other data collections of federal agencies. The Federal Education Act of 1974 contained an amendment by Senator Buckley, popularly labelled the "Buckley Amendment," that forced all educational institutions that receive federal grants to make all records available as noted above. The problems created by the inclusion of medical reports led to a second amendment to the law, restricting disclosure of medical records on demand of advanced school students to a physician of data-subject's choice. Schools through high school are still vulnerable. Some school officials have begun to add a warning about this to their requests for medical information from treating physicians of the school child involved. A number of additional bills in Congress and in the California legislature include such a disclosure section. A few include the provision of limiting the disclosure to a physician of the subject's choice.

The disclosure legislation is based on the general proposition that patients will fare better if well informed about the whole medical process involving them. Those who advocate this usually suggest that opposition is for the purpose of hiding poor medical practice (24, 25). They overlook the fact that there is some medical data that is intolerable for some patients and can cause emotional damage. It is not uncommon to find patients blocking and practicing denial about facts that they themselves have disclosed at another time. Sometimes the true state and diagnosis is intolerable and forced disclosure to the patient proves destructive (11). Disclosure can be "forced" when the family insists the patient get a copy of the record. It is "forced" when the patient wants to "prove" that the condition is other than suggested by partial knowledge that must be revealed for other reasons. Such experiences create doubts about the ultimate wisdom of a blanket disclosure of records to all patients. Such records often contain confidential information about parties other than the patient. They, too, need protection of their confidential disclosures. The suggestion of having a physician of the patient's choice monitor the information is but a compromise, to shift the burden of protecting the patients from themselves—a not infrequent need in medical care.

WORLDWIDE ACTIVITY SEEKING SOLUTIONS

There are enough complexities to demonstrate that there can be no simple resolution of "the right of privacy" vs "the right to know." This is reflected in increasing numbers of conferences on the subject. The World Medical Association Congress of October, 1973 has been mentioned. Most often they lead to generalized resolutions that do not alter the situation, but do lead to exchange of information about conflicting needs. Congressional hearings (22a, b) have been extensive and led to the Privacy Act of 1974. This is limited to Federal records because broader complexities were too great. It did include a provision for a Privacy Protection Study Commission to explore for further legislative relief. This commission has been holding formal hearings all over the country to obtain facts from all segments of society, not overlooking the problem of protecting medical records per se. The National Bureau of Standards authorized a project to study the role of computers and needs for protecting medical records. This study, directed by Alan F. Westin, is in the process of being reported after two years of detailed study and consultation. Earlier studies were commissioned by the Secretary of Health, Education, and Welfare (HEW), (22b) and the National Academy of Science (22c). They had a profound effect on the content of the Privacy Act of 1974. Confirming the fact that the concern is worldwide is the reference to similar studies in other countries in the bibliography of Ware's HEW report and in Westin's study.

ROLE OF MEDICAL ORGANIZATIONS

Organized medicine has also become concerned. The California Medical Association took the lead in calling attention to the problem in 1970 and trying to mobilize the American Medical Association (AMA) to take definitive action for protecting the rights of patients. As a result the AMA House of Delegates supported and recommended action for the protection of confidentiality in a great number of resolutions from 1972 through 1976. However, the positive achievement has been repeated formulation of a model state law. Another such model, considered by some to have basic flaws due to the complexity of its global approach, was approved at the 1976 annual meeting. However, when the AMA Council on Medical Services reviewed a standard insurance claim form in March, 1973, they did not respond favorably to recommendations that the AMA withdraw approval of the unlimited authorization for release of medical information. Persuaded by the insurance industry representatives to accept a reworded form of equal legal comprehensiveness, the current form, approved by the House of Delegates in 1973, contains this authorization with the imprinted approval of the AMA.[3]

When the House of Representatives was considering a new Federal Code of Evidence, containing Rule 504, extending privilege to the psychotherapeutic medical data and communications between all physicians and their patients, the former Executive Vice President of the AMA, Ernest R. Howard, wrote Congressman Rodino January 31, 1973, opposing Rule 504 and wanting a physician-patient privilege that is legally meaningless, as discussed above. He wrote a similar letter to Congressman Hungate on August 1, 1973. On August 26, 1974, after the CMA and AMA House of Delegates protest, Dr. Howard retracted opposition to Rule 504, but by then the damage was done. We now have no federal privilege protection for medical matter. We are reduced to state codes of meaningless physician-patient privilege in 32 states; psychotherapist-patient privilege restricted to psychiatrists and psychologists in seven states; and none for other physicians in their psychotherapeutic dealings with their patients.

---

[3] At a Federal Trade Commission quasi-judicial hearing in San Francisco on February 4, 1976, attorneys for Retail Corporation (Equifax) used this to prove that the AMA approved of the practices of the insurance industry and the Retail Credit Corporation in their wide collecting of medical records and information.

The American Psychiatric Association (APA) became alerted to the problem in 1970. Its Task Force foresaw the greater hazards in National Health Insurance, since all proposed legislation gave no attention to the medical-record protection that would be needed. The APA proposed this need in 1972 (26). The recognition that it is more than a physician's problem led to the recommendation that the public and other professionals work together to seek solutions for problems resulting from the conflicting interests. This resulted in the Key Biscayne Conference alluded to above (22d, 23), attended by representatives from over fifty national organizations—medical, legal, educational, religious, consumer groups, computer experts, the press, the insurance industry, and the government, including the military (27). The conference strongly supported the foundation of a private multi-discipline organization to keep abreast of the developing problems and efforts to deal with them. The results of the Key Biscayne Conference led to the recognition that there should be an ongoing dialogue, kept current, between conflicting viewpoints if constructive compromises are to be achieved. This National Commission on the Confidentiality of Health Records was incorporated early in 1976 with 17 national organizations as charter members. Although the AMA participated in all the planning meetings for the Key Biscayne Conference and participated in that conference, again it did not follow up on the resolutions of activity for confidentiality, electing not to join in the National Commission.[4]

The current ferment suggests that there will be much change in the future for medical records and their confidentiality. Consumer concern is reflected in the multiple studies that have legislative goals. It will affect record keeping in hospitals, clinics, and physicians' offices. Teaching and research, as well as treatment, will be affected. It will certainly affect the welfare of patients and, thereby, medical practice. The role of physicians in this ferment and its consequences is unclear beyond their majority affirmation, "I am in favor of confidentiality." Solutions will not be as simple as those for the problems faced by the Hippocratic group.

[4] Correspondence with James A. Sammons, AMA Executive Vice President, failed to elicit the reasons, but other sources indicated that the AMA wished to maintain an "independent stand." There does not seem to be any recognition of the growing dissatisfaction of patients with the current situation.

Literature Cited

1. Edelstein, L. 1943. The Hippocratic Oath. *Bull. Hist. Med.,* ed. H. E. Sigeret, 7:Suppl. 1. Baltimore: Johns Hopkins Press. 64 pp.

2. Pearson, H. L. Jr. 1958. AMA Principles of Medical Ethics. *J. Am. Med. Assoc.,* Spec. ed., Sect. 9, p. 49. 59 pp.

3a. Wigmore, J. H. 1961. *Evidence in Trials at Common Law,* 8:Sect. 2380, p. 818. Boston: Little, Brown. 10 vol. Revised ed.

3b. Slovenko, R. 1973. *Psychiatry and Law,* p. 62. Boston: Little, Brown. 726 pp.

4. Watson, A. S. 1960. *Confidentiality and Privileged Communication in the Practice of Psychiatry,* Rep. No. 45, pp. 88–113. New York: Group Adv. Psychiatry. Revised ed.

5a. Committee on Rules of Practice and Procedure, Judicial Conference of the U.S. 1969. *Preliminary Draft of Proposed Rules of Evidence for the U.S. District Courts and Magistrates,* footnote. Rule 504, p. 53

5b. Slovenko, R. see Ref. 3b, pp. 62–63

6a. Wigmore, J. H. 1961. see Ref. 3a, Sect. 2285, p. 527

6b. Grossman, M., Barchas, J., Johnson, R. G., Satten, J. 1975. Confidentiality and third parties. In *Task Force Report 9,* p. 44. Washington DC: Am. Psychiatr. Assoc. 59 pp.

7a. *In re Lifschutz.* 1970. 2Cal3d at 430, 467P.2d at 566, 85 Cal Rptr. at 838

7b. *California Code of Evidence.* 1966. Section 912, p. 491. St. Paul, Minn: West

7c. Louisell, D. W., Sinclair, K. Jr. 1971. The Supreme Court of California 1969–1970. *Calif. Law Rev.* 59:41

8. Grossman, M. 1971. Insurance reports as a threat to confidentiality. *Am. J. Psychiatry* 128:64–68

9. Porter, S. 1976. Mail order insurance pitfalls. *San Francisco Chron.,* p. 50 (May 3)

10. Tooley, D. 1976. Testimony of Federal Privacy Protection Study Commission. *Los Angeles Times.* Pt. II, p. 1 (June 12)

11. Grossman, M. See Ref. 6, Append. H, p. 53

12. Testimony before Federal Privacy Protection Study Commission, Los Angeles, June 11, 1976

13. Grossman, M. 1975. See Ref. 6 Append. B, pp. 14–20

14a. *In re Lifschutz.* 1970. 2Cal.3d at 431–32, 467 P.2dt at 567, 85Cal. Rptr. at 839

14b. Committee on Rules of Practice & Procedures 1969. See Ref. 5a

14c. Louisell, D. W., Sinclair, K. Jr. 1971. See Ref. 7c, p. 42

15. Grossman, M. 1975. The psychiatrist and the subpoena. *Bull. Am. Acad. Psychiatry Law* 1:245–54

16. Solomon, P., Kleeman, S. T., Curran, W. J. 1971. Confidentiality in psychiatric screening for security clearance. *Am. J. Psychiatry.* 127:1566–68

17a. Grossman, M., Satten, J. 1975. See Ref. 6, Append. F, p. 34

17b. Satten, J. et al 1970. *The Field of Family Therapy,* Rep. No. 78. New York: Group Adv. Psychiatry

18a. Slovenko, R. 1973. See Ref. 3b, p. 500

18b. Peer, I. N. 1976. Confidentiality and consent in psychiatric treatment of minors. *J. Leg. Med.* 4:12–16

18c. Modlin, H. C. 1976. The discharged mental patient in the community. *Psychiatry Dig.* 37:16–21

19. *Tarasoff v. Regents of the Univ. of Calif.* 118 Cal. Rptr. 129 (1974) Also S.F.23042, Super. Ct. 405694 (rehearing) July 1, 1976

20. California Code of Evidence. 1966. See Ref. 7b, Sect. 1024, p. 630

21. Petris, N. C. 1972. *The Dilemma of Mental Commitments in California,* p. 143. Sacramento: Calif. Legislature Assem. 204 pp.

22a. Ervin, S. J. Jr. 1974. *Protecting Individual Privacy in Federal Gathering, Use, and Disclosure of Information.* Rep. Comm. Gov. Oper., US Senate, accompanying S. 3418, 93rd Congr., 2nd Sess. Washington DC: GPO

22b. Ware, W. 1973. *Records, Computers and the Rights of Citizens.* Dept. HEW Publ. No. (OS) 73–97 xi. 346 pp. Washington DC: GPO

22c. Westin, A. F., Baker, M. A. 1972. *Data Banks in A Free Society.* Natl. Acad. Sci. New York: Quadrangle. 522 pp.

22d. Participants. 1974. Conf. Confidentiality of Health Rec., Key Biscayne, Fla., Nov. 6–9.

23. Spingarn, N. D. 1975. *Confidentiality,* pp. 17–18. Washington DC: Am. Psychiatr. Assoc. 58 pp.

24. Spingarn, N. D. See Ref. 23, p. 14

25. Weed, L. L. 1974. The public needs *must* be met. *Prism* 2:22–26

26. Grossman, M. 1975. See Ref. 6, Append. C, p. 20; Append. F, p. 35

27. Spingarn, N. D. See Ref. 23, pp. 43–52

Ann. Rev. Med. 1977. 28:57–65
Copyright © 1977 by Annual Reviews Inc. All rights reserved

IMPAIRED PERIPHERAL CONVERSION OF THYROXINE TO TRIIODOTHYRONINE

❖7222

Ralph R. Cavalieri, M.D.
Nuclear Medicine Service and Endocrine Section, Veterans Administration Hospital;
and Departments of Medicine and Radiology, University of California, San Francisco,
California 94121

Basil Rapoport, M.B., Ch.B.
Endocrine Section, Veterans Administration Hospital;
and Department of Medicine, University of California, San Francisco, California 94121

THYROID HORMONE METABOLISM IN MAN

The thyroid gland of healthy adults secretes approximately 78 μg (100 nmole) thyroxine (T_4) per day. Of this, about 80 nmole undergo deiodination in extrathyroidal sites. The deiodinative metabolism of T_4 probably occurs in a stepwise manner. Removal of one iodine from the outer (phenolic) ring of T_4 (5'deiodination) yields 3,5,3'-triiodothyronine (T_3); removal of an inner-(tyrosyl) ring iodine (5-deiodination) yields 3,3', 5'-T_3 [reverse-T_3 (rT_3)]. Non-deiodinative pathways include excretion of T_4 and its conjugates (glucuronide, sulfate) in the feces and oxidative deamination to the acetic-acid derivative [tetraiodothyroacetic acid (tetrac)]. Together, these routes account for nearly all of the remaining 20 nmole of T_4 turned over each day.

The peripheral conversion of T_4 to T_3 has been demonstrated in humans (1–3). Estimates of the extent of such conversion, i.e. the conversion ratio (CR), obtained by different methods, indicate that in euthyroid individuals about 35% of T_4 turnover (35 nmole) proceeds via the generation of T_3 (2–4).

Recent studies have indicated that rT_3 is also a major product of T_4 metabolism in humans, accounting for about 45 of the 80 nmole of T_4 undergoing deiodination each day (5, 6). Thus the formation of T_3 and rT_3 together account for all of the T_4 deiodinated (Figure 1). Subsequent deiodination of T_3 and rT_3 produce various diiodo- and monoiodothyronines (7, 8).

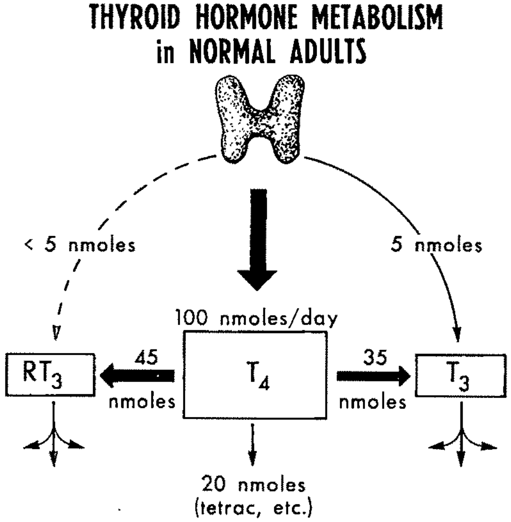


Figure 1 Schematic representation of major pathways of thyroxine metabolism in normal human adults. Rates are expressed in nmole (24 hr)$^{-1}$ and ought to be considered as approximations, based upon available data. T_4, thyroxine; T_3, 3,4,3'-triiodothyronine; RT_3, 3,3'5'-triiodothyronine (reverse-T_3); tetrac, acetic-acid derivative of T_4.

T_3 is several times more potent than T_4 itself; thus conversion to T_3 represents a process of hormonal activation. None of the other deiodinated derivates of T_4 exhibit calorigenic activity in the intact organism (9). Reverse-T_3 has been shown to antagonize the effects of T_4 in animals, but only in high doses (10). The active form of the thyroid hormone has been discussed in a previous volume of this series (11).

Sources of T_3 in Man

The production rate of T_3 is approximately 26 μg (40 nmole) per day (4, 6, 12). Since about 35 nmole of T_3 are derived by peripheral 5'-deiodination of T_4 in euthyroid individuals, this leaves only 5 nmole for direct secretion of T_3 from the thyroid. These estimates, based upon average calculated values for T_3 turnover rates and CR reported from several laboratories, are subject to revision. Nevertheless, there is general agreement that the vast bulk of circulating T_3 is generated from T_4 by peripheral conversion and only a small proportion is derived from the gland directly. In contrast to euthyroid individuals, in untreated hyperthyroid patients the thyroid gland secretes a larger proportion of the total T_3 produced. As a corollary, a decrease in circulating T_3 levels in the face of a constant serum T_4 concentration in a euthyroid individual is likely to represent a decline in the rate of peripheral conversion of T_4 to T_3 rather than a change in the rate of thyroidal secretion of T_3, assuming that the T_3 clearance rate is not selectively increased.

Biochemical Mechanism of T_4 to T_3 Conversion

5'-deiodination of T_4 has been demonstrated in tissue slices (13, 14), cell cultures (15, 16), and cell-free homogenates (17–20) of a number of tissues. Although the evidence at hand indicates an enzymatic mechanism, the enzyme has not been isolated, and its tissue distribution and intracellular localization are still uncertain. Even less is known about the formation of rT_3 from T_4 (5-deiodination), which is presumably mediated by a different enzyme than the 5'-deiodinase. Reverse-T_3 has been shown to inhibit competitively the formation of T_3 from T_4 in liver homogenates (20) and kidney slices (21). Although it is uncertain whether in vivo concentrations of rT_3 ever reach levels sufficiently high to effect this inhibition, these observations raise the intriguing possibility that the product of an inactivating pathway of T_4 metabolism may modulate the rate of the activating pathway. This could represent one of many possible mechanisms of regulatory control over the reaction by which T_3 is formed within tissues.

T_4 to T_3 CONVERSION IN DIVERSE CLINICAL CONDITIONS

Normal Human Neonates

The serum concentration of T_3 is lower (22), and that of rT_3 higher (23) in umbilical cord blood than in the blood of adults. The same pattern has been found in amniotic fluid of normal human pregnancy (24). Studies of T_3 and rT_3 kinetics in fetal and adult sheep indicate that impairment of T_4 to T_3 conversion is primarily responsible for the low T_3 in the fetus and newborn and that T_4 to rT_3 conversion is relatively active in the fetus (25). Both in humans (22) and in sheep (25), serum T_3 levels increase within hours after birth, but rT_3 concentrations fall to the normal adult level only after several days (23, 25). In the human at least, a portion of the early postnatal recovery in T_3 levels seems due to direct thyroidal secretion probably responding to the TSH -surge (26). Clearly the pattern of T_4 metabolism in the fetus and newborn is different from that in the human adult.

Effect of Age

The serum concentration of T_3, but not of T_4, declines with advancing age (27). Using the turnover method, Wenzel & Horn have reported a 20% lower CR in healthy men, aged 73 to 82, compared to a group of 21- to 29-year-olds, a finding that seemed to account for most of the age-dependent decrease in the serum T_3 level (28).

Thyroid Disease

Based upon the sparse data available to date, it appears that CR is normal or only slightly elevated in untreated hyperthyroidism (29). The absolute T_3 production rate is increased, as much as seven times normal (12), so the major portion of the excess T_3 produced in this condition can be accounted for by direct secretion from the thyroid.

In untreated hypothyroid patients, CR was found to be nearly twice normal in one study (30). Others have reported an increased CR in hypothyroidism, but to a smaller degree (29). Even when CR is increased, however, the absolute amount of T_3 derived from T_4 is small, since the absolute turnover of T_4 is below normal in hypothyroid individuals (12).

Caloric Restriction

Fasting leads to a prompt (within 24 hours) fall in serum T_3 concentration in obese patients (31, 32), as well as in subjects of normal body weight (33). Total serum T_4 remains relatively unchanged, but free T_4 concentration increases slightly (31). The same phenomenon has been observed in protein-calorie malnutrition (34, 35), and anorexia nervosa (36, 37). Re-feeding restores serum T_3 to normal (34, 35). The decline in serum T_3 is sensitive to the carbohydrate in the diet (32, 38). That the changes in serum T_3 levels are due to alterations in T_4 to T_3 CR is supported by the finding of similar diet-related changes in subjects receiving L-T_4 in suppressive doses (39). Increases and decreases in serum rT_3 concentration with fasting and re-feeding, respectively, have been found to coincide with reciprocal changes in serum T_3 (38, 39), suggesting that caloric deprivation, especially of carbohydrate, causes a diversion in T_4 metabolism away from the calorigenically active product, T_3, toward the much less active derivative, rT_3.

Hepatic Disease

Patients with hepatic cirrhosis often have diminished, sometimes profoundly low serum levels of total and free T_3 (40, 41) and normal or slightly elevated total and free T_4 (41, 42). The findings most likely reflect impaired T_4 to T_3 conversion. Nomura et al have in fact estimated CR by the turnover method and found an average of 15.6% ± 6.6 [standard deviation (S.D.)] in four patients with advanced cirrhosis, compared with 35.7% (± 3.3) in 7 normal subjects (43). Improvement in hepatic function is associated with a tendency for serum T_3 levels to return toward normal (43). Thus the liver seems to play a dominant role in the peripheral metabolism of T_4, particularly in the generation of T_3.

Other Systemic Disorders

The combination of a decreased serum T_3 concentration and normal or slightly elevated serum T_4 has been reported in a wide variety of systemic illnesses besides hepatic disease. These conditions include disseminated malignancy (40, 44), renal failure (40, 45), febrile states (46), and postsurgical states (47). Bermudez et al (44) reported that 70% of a group of clinically euthyroid patients with serious systemic illness exhibited an abnormally low total serum T_3. In about one half of these, the low T_3 level was attributable to decreased plasma binding, free-T_3 concentration being normal. In the remainder (32% of the entire group), the free as well as total T_3 level was decreased. Tissue concentrations of T_3 have been measured at autopsy and found to be decreased in patients who died after prolonged illness, compared to victims of sudden accidents (48). It is now recognized that a significant proportion of patients with systemic illness have low circulating levels of total and free T_3 and normal or high free-T_4 levels, and that in most if not all cases, this reflects impaired peripheral conversion of T_4 to T_3.

Reverse-T_3 in Systemic Illness

Serum rT_3 levels are elevated in some but not all patients in whom serum T_3 concentrations are decreased (46, 49–51). Cases of cirrhosis tend to show the lowest serum T_3 values. However, it is not certain whether the elevated rT_3 concentrations reflect increased production from T_4 or decreased rT_3 clearance. Flock has shown in the dog that hepatectomy severely impairs the production of T_3, but not of rT_3 from T_4 (7). This finding and the observations in patients with liver disease would suggest that normal conversion of T_4 to T_3 requires intact liver function and that the production of rT_3 occurs in extrahepatic tissues.

Effects of Drugs

Propylthiouracil (PTU), but not methimazole, inhibits the peripheral deiodination of T_4 in rats (52) and in humans (53), and diminishes the biologic effectiveness of exogenous T_4 (54). Recent studies specifically directed at T_4-to-T_3 conversion have shown that PTU inhibits 5'-deiodination in man (55, 56). Although the block is not complete, this peripheral effect of PTU offers an added therapeutic advantage in the treatment of severe thyrotoxicosis over and above its effect on the thyroid gland (57).

Acute administration of the glucocorticoid dexamethasone in total daily doses of 8 mg causes a prompt fall in serum T_3 (58, 59) and an elevation in serum rT_3 (59) in normal individuals, patients with hyperthyroidism, and hypothyroid patients receiving T_4-replacement therapy. Although the magnitude and duration of the changes in circulating T_3 and rT_3 differ somewhat among the various groups, the pattern suggested an effect of dexamethasone on the monodeiodination of T_4, away from T_3 and toward rT_3 production.

METABOLIC STATUS OF PATIENTS WITH IMPAIRED T_4 TO T_3 CONVERSION

Because T_3 accounts for a major portion of the total metabolic effect of thyroid hormone, there is a question as to whether individuals with low serum T_3 due to impaired conversion from T_4 are euthyroid or hypothyroid. One of the reasons for confusion is the paucity of sensitive and specific tests to quantitate thyroid-hormone action at the cellular level.

The most sensitive indicator of thyroid-hormone action is provided by the serum TSH concentration. In the absence of hypothalamic and pituitary disease, the synthesis and secretion of TSH by the pituitary are exquisitely sensitive to alterations in available thyroid hormone. Primary hypothyroidism is associated with an elevated basal serum TSH concentration and an augmented TSH response to administration of thyrotropin-releasing hormone TRH (60).

In the majority of individuals with low serum T_3 associated with nonthyroidal systemic illness (40, 50), fasting (31, 33, 61), or protein-calorie malnutrition (34, 35), serum TSH is within the normal range. In two studies of patients with cirrhosis, the mean serum TSH levels were slightly higher than the normal mean values (but still within the normal range) (41, 43). In one of these groups the TSH response to TRH was, however, no different than in controls (41). In another report, a group of patients with diverse systemic illness showed normal basal TSH values and TSH responses to TRH, although the latter were slightly exaggerated (44). The evidence available to date indicates that the pituitary thyrotrophs seem to sense a sufficiency of thyroid hormone. For this reason, the majority opinion is that these individuals are euthyroid.

It is possible that these patients with a subnormal serum (total and free) T_3 concentration are maintained in an euthyroid state by circulating T_4. In some but not all of those cases in whom the measurements were done, serum free-T_4 concentrations were variably elevated (31, 34, 41–44). There are good reasons to believe than T_4 possesses significant intrinsic hormonal activity without the need for conversion to T_3 (11). It is conceivable that increases in circulating free T_4 levels, perhaps even more marked than present in vitro measurements would indicate, tend to compensate for the subnormal free-T_3 concentrations, at least in some cases.

Apparent euthyroidism, both clinically and by measurement of TSH, in the face of a low serum T_3 also makes it necessary to reexamine the widely held concept that each individual has a constant requirement for thyroid hormone. During the stress of systemic illness or during restricted caloric intake, a decline in circulating T_3 levels may be an appropriate response that has protective value to the organism. Oxygen consumption and respiratory quotient decline during caloric restriction in obese subjects, and can be restored by administration of T_3 (62). A hypometabolic state offers obvious advantages during starvation and may confer some protection against the excessive catabolism associated with systemic illness. Viewed in this light, the inhibition of T_4 to T_3 conversion may be an appropriate response under certain conditions.

However, it still cannot be excluded that individuals with nonthyroidal systemic illness and low serum-T_3 levels are in fact hypothyroid, despite the lack of a clearly elevated serum-TSH level. First, hypothalamic-pituitary dysfunction is a possibility and has been implicated in anorexia nervosa (36). Second, the sensitivity of the pituitary thyrotroph may be increased, so that a subnormal level of serum T_3 is sufficient to prevent the expected rise in serum TSH. The observed normal TSH responses to administered TRH in most of these patients would be consistent with such an alteration in sensitivity. Against this possibility is the fact that the serum free-T_4 concentration is normal or somewhat elevated in patients with nonthyroidal systemic illness (41–44) or during starvation (32, 34). In patients with secondary or tertiary (hypothalamic) hypothyroidism, serum T_4 typically is low.

For practical purposes, most would agree that in a patient with a systemic illness an abnormally low serum-T_3 concentration (total or free) does not per se signify hypothyroidism, at least the type requiring replacement therapy. It would seem appropriate to demand such evidence of diminished thyroid-gland function as a decreased serum free-T_4 concentration or an elevated serum TSH concentration before considering replacement with thyroid hormone.

Literature Cited

1. Braverman, L. E., Ingbar, S. H., Sterling, K. 1970. Conversion of thyroxine (T_4) to triiodothyroxine (T_3) in athyreotic human subjects. *J. Clin. Invest.* 49:855–64
2. Pittman, C. S., Chambers, J. B., Read, V. H. 1971. The extrathyroidal conversion rate of thyroxine to triiodothyronine in normal man. *J. Clin. Invest.* 50:1187–96
3. Surks, M. I., Schadlow, A. R., Stock, J. M., Oppenheimer, J. H. 1973. Determination of iodothyronine absorption and conversion of L-thyroxine (T_4) to L-triiodothyronine (T_3) using turnover rate techniques. *J. Clin. Invest.* 52:805–11
4. Braverman, L. E., Vagenakis, A., Downs, P., Foster, A. E., Sterling, K., Ingbar, S. H. 1973. Effects of replacement doses of sodium-L-thyroxine on the peripheral metabolism of thyroxine and triiodothyronine in man. *J. Clin. Invest.* 52:1010–17
5. Chopra, I. J., 1976. An assessment of daily turnover and significance of thyroidal secretion of 3,3',5'-triiodothyronine (reverse T_3) in man. *J. Clin. Invest.* 58:32–40
6. Gavin, L., Castle, J., McMahon, F., Martin, P., Hammond, M., Cavalieri, R. R. 1976. Metabolic clearance and production rates of 3,5,3'-triiodothyronine (T_3) and 3,3',5'-T_3 (reverse T_3) in man. *Clin. Res.* 24:427A (Abstr.)
7. Flock, E. V., Bollman, J. L., Grindlay, J. H., Stobie, G. H. 1961. Partial deiodination of L-thyroxine. *Endocrinology* 69:626–37
8. Rudolph, M., Sakurada, T., Vagenakis, A., Fang, S., Braverman, L., Ingbar, S. H. 1976. Demonstration of sequential monodeiodination as the major pathway of iodothyronine metabolism. *Clin. Res.* 429A (Abstr.)
9. Pittman, C. S., Pittman, J. A. 1974. Relation of chemical structure to the action and metabolism of thyroactive substances. *Handb. Physiol.* 3:233–53
10. Pittman, C. S., Barker, S. B. 1959. Antithyroxine effects of some thyroxine analogues. *Am. J. Physiol.* 197:1271–74
11. Ingbar, S. H., Braverman, L. 1975. Active form of the thyroid hormone. *Ann. Rev. Med.* 26:443–49
12. Nicoloff, J. T., Low, J. C., Dussault, J. H., Fisher, D. A. 1972. Simultaneous measurement of thyroxine and triiodothyronine peripheral turnover kinetics in man. *J. Clin. Invest.* 51:473–83
13. Albright, E. C., Larson, F. C. 1959. Metabolism of L-thyroxine by human tissue slices. *J. Clin. Invest.* 38:1899–1903
14. Green, W. L., 1976. Thyroxine metabolism by rat liver slices: evidence for a specific T_3-forming pathway. In *Thyroid Research,* ed. J. Robbins, L. E. Braverman, pp. 239–44. Amsterdam: Excerpta Medica. 638 pp.
15. Sterling, K., Brenner, M. A., Saldanha, V. F. 1972. Conversion of thyroxine to triiodothyronine by cultured human cells. *Science* 179:1000–1
16. Refetoff, S., Matalon, R., Bigazzi, M. 1972. Metabolism of L-thyroxine (T_4) and L-triiodothyronine (T_3) by human fibroblasts in tissue culture: Evidence for cellular binding proteins and conversion of T_4 to T_3. *Endocrinology* 91:934–47
17. Hesch, R. D., Brunner, G., Soling, H. D. 1975. Conversion of thyroxine (T_4) and triiodothyronine (T_3) and the subcellular localization of the converting enzyme. *Clin. Chim. Acta* 59:209–13
18. Visser, T. J., Does-Tobe, I.v.d., Doctor, R., Henneman, G. 1976. Conversion of thyroxine into triiodothyronine by rat liver homogenate. See Ref. 14, pp. 235–38
19. Chiraseveenuprapund, P., Buergi, U., Goswami, A., Rosenberg, I. N. 1976. Formation of triiodothyronine from L-thyroxine in rat liver homogenate. See Ref. 14, pp. 244–47
20. Chopra, I. J. 1976. Study of extrathyroidal conversion of T_4 to T_3 in vitro: evidence that reverse T_3 is a potent inhibitor of T_3 production. *Clin. Res.* 24:142A (Abstr.)
21. Larson, F. C., Albright, E. C. 1961. Inhibition of L-thyroxine monodeiodination by thyroxine analogs. *J. Clin. Invest.* 40:1132–38
22. Abuid, J., Stinson, A., Larsen, P. R. 1973. Serum triiodothyronine and thyroxine in the neonate and the acute increases in these hormones following delivery. *J. Clin. Invest.* 52:1195–99
23. Chopra, I. J., Sack, J., Fisher, D. A. 1975. Circulating 3,3',5'-triiodothyronine (reverse T_3) in the human newborn. *J. Clin. Invest.* 55:1137–41
24. Chopra, I. J., Crandall, B. F. 1975. Thyroid hormones and thyrotropin in amniotic fluid. *N. Engl. J. Med.* 293:740–43
25. Chopra, I. J., Sack, J., Fisher, D. A. 1975. 3,3',5'-triiodothyronine (reverse T_3) and 3,3',5'-triiodothyronine (T_3) in

fetal and adult sheep: studies of the metabolic clearance rates, production rates, serum binding, and thyroidal content relative to thyroxine. *Endocrinology* 98:1080–88

26. Fisher, D. A., Odell, W. D. 1969. Acute release of thyrotropin in the newborn. *J. Clin. Invest.* 48:1670–77
27. Rubenstein, H. A., Butler, V. P., Butler, J., Werner, S. C. 1973. Progressive decrease in serum triiodothyronine concentrations with human aging: radioimmunoassay following extraction of serum. *J. Clin. Endocrinol. Metab.* 37:247–53
28. Wenzel, K. W., Horn, W. R. 1976. Triiodothyronine (T_3) and thyroxine (T_4) kinetics in aged man. See Ref. 14, pp. 270–73
29. Maquire, S. B., Denneby, A., Cullen, M. J. 1976. The effect of thyrotoxicosis and hypothyroidism on the extrathyroidal conversion of thyroxine to triiodothyronine in man. See Ref. 14, pp. 259–62
30. Inada, M., Kasagi, K., Jurata, S., Kazama, Y., Takayama, H., Torizuka, K., Fukase, M., Soma, T. 1975. Estimation of thyroxine and triiodothyronine distribution and of the conversion rate of thyroxine to triiodothyronine in man. *J. Clin. Invest.* 55:1337–48
31. Portnay, G. E., O'Brian, J. T., Bush, J., Vagenakis, A. G., Azizi, F., Arky, R. A., 1974. The effect of starvation on the concentration and binding of thyroxine and triiodothyronine in serum and on the response to TRH. *J. Clin. Endocrinol. Metab.* 39:191–94
32. Spaulding, S. W., Chopra, I. J., Sherwin, R. S., Lyall, S. S. 1976. Effect of caloric restriction and dietary composition on serum T_3 and reverse T_3 in man. *J. Clin. Endocrinol. Metab.* 42:197–200
33. Merimee, T. J., Fineberg, E. S. 1976. Starvation-induced alterations of circulating thyroid hormone concentrations in man. *Metabolism* 25:79–83
34. Chopra, I. J., Smith, S. R. 1975. Circulating thyroid hormones and thyrotropin in adult patients with protein-calorie malnutrition. *J. Clin. Endocrinol. Metab.* 40:221–27
35. Ingenbleek, Y., Beckers, C. 1975. Triiodothyronine and thyroid stimulating hormone in protein calorie malnutrition in infants. *Lancet* 2:845–47
36. Miyai, K., Yamamoto, T., Azukizawa, M., Ishibashi, K., Kumahara, Y. 1975. Serum thyroid hormones and thyrotropin in anorexia nervosa. *J. Clin. Endocrinol. Metab.* 40:334–38
37. Moshang, T., Parks, J. S., Baker, L., Vaidya, V., Utiger, R. D. 1975. Low serum triiodothyronine in patients with anorexia nervosa. *J. Clin. Endocrinol. Metab.* 40:470–73
38. Danforth, E. Jr., Tyzbir, E. D., Horton, E. S., Sims, E. A. H., Burger, A. G., Braverman, L. E., Vagenakis, A. G., Ingbar, S. H. 1976. Reciprocal changes in serum triiodothyronine (T_3) and reverse (rT_3) induced by altering the carbohydrate content of the diet. *Clin. Res.* 24:271A (Abstr.)
39. Vagenakis, A. G., Burger, A., Portnay, G. I., Rudolph, M., O'Brian, J. T., Azizi, F., Arky, R. A., Nicod, P., Ingbar, S. H., Braverman, L. E. 1975. Diversion of peripheral thyroxine metabolism from activating to inactivating pathways during complete fasting. *J. Clin. Endocrinol. Metab.* 41:191–94
40. Carter, J. N., Corcoran, J. M., Eastman, C. J., Lazarus, L. 1974. Effect of severe, chronic illness on thyroid function. *Lancet* 27:971–74
41. Chopra, I. J., Solomon, D. H., Chopra, V., Young, R. T., Chua Teco, G. N. 1974. Alterations in circulating thyroid hormones and thyrotropin in hepatic cirrhosis: evidence for euthyroidism despite subnormal serum triiodothyronine. *J. Clin. Endocrinol. Metab.* 39:501–11
42. Inada, M., Sterling, K. 1967. Thyroxine turnover and transport in Laennec's cirrhosis of the liver. *J. Clin. Invest.* 46:1275–82
43. Nomura, S., Pittman, C. S., Chambers, J. B. Jr., Buck, M. W., Shimizu, T. 1975. Reduced peripheral conversion of thyroxine to triiodothyronine in patients with hepatic cirrhosis. *J. Clin. Invest.* 56:643–52
44. Bermudez, F., Surks, M. I., Oppenheimer, J. H. 1975. High incidence of decreased serum triiodothyronine concentration in patients with nonthyroidal disease. *J. Clin. Endocrinol. Metab.* 41:27–40
45. Fang, V. S., Lim, V. S., Refetoff, S. 1974. Triiodothyronine (T_3) hypothyroidism in patients with renal insufficiency. Program Ann. Meet. Am. Thyroid Assoc., 50th, p. T-16 (Abstr.)
46. Chopra, I. J., Chopra, V., Smith, S. R., Reza, M., Solomon, D. H. 1975. Reciprocal changes in serum concentrations of 3,3',5'-triiodothyronine (reverse T_3) and 3,3',5-triiodothyronine (T_3) in sys-

temic illness. *J. Clin. Endocrinol. Metab.* 41:1043–49

47. Burr, W. A., Griffiths, R. S., Black, E. G., Hoffenberg, R., Meinhold, H., Weinzel, K. W., 1975. Serum triiodothyronine and reverse triiodothyronine concentrations after surgical operation. *Lancet* 2:1277–79
48. Sullivan, P. R. C., Bollinger, J. A., Reichlin, S. 1973. Selective deficiency of tissue triiodothyronine: a proposed mechanism of elevated free thyroxine in the euthyroid sick. *J. Clin. Invest.* 52:83a
49. Nicod, P., Burger, A., Staeheli, V., Vallotton, M. B. 1976. A radio-immunoassay for 3,3',5'-triiodo-L-thyronine in unextracted serum; method and clinical results. *J. Clin. Endocrinol. Metab.* 42:823–29
50. Burger, A., Nicod, P., Suter, P., Vallotton, M. D., Vagenakis, A., Braverman, L. 1976. Reduced active thyroid hormone levels in acute illness. *Lancet* 1:653–55
51. Gavin, L., Rapoport, B., Hammond, M., Cavalieri, R. R. 1976. Variable serum reverse-T_3 (RT_3) and decreased T_3 concentrations in non-thyroidal systemic illness (SI). *Clin. Res.* 24: 272A (Abstr.)
52. Escobar de Rey, F., Morreale de Escobar, G. 1961. The effect of propylthiouracil, methylthiouracil and thiouracil on the peripheral metabolism of L-thyroxine in thyroidectomized, L-thyroxine maintained rats. *Endocrinology* 69:456–65
53. Hershman, J. M. 1974. Effect of 5- and 6-propylthiouracil on the metabolism of L-thyroxine in man. *J. Clin. Endocrinol. Metab.* 24:173–79
54. Barker, S. B., Kiely, C. E. Jr., Lipner, H. J. 1949. Metabolic effects of thyroxine injected into normal, thiouracil-treated and thyroidectomized rats. *Endocrinology* 45:624–27
55. Saberi, M., Sterling, F., Utiger, R. 1975. Reduction in extrathyroidal triiodothyronine production by propylthiouracil in man. *J. Clin. Invest.* 55: 218–23
56. Geffner, D. L., Azukizawa, M., Hershman, J. M. 1975. Propylthiouracil blocks extrathyroidal conversion of thyroxine to triiodothyronine and augments thyrotropin secretion in man. *J. Clin. Invest.* 55:224–29
57. Abuid, J. Larsen, P. R. 1974. Triiodothyronine and thyroxine in hyperthyroidism. *J. Clin. Invest.* 54:201–8
58. Duick, D. S., Warren, D. W., Nicoloff, J. T., Otis, C. L., Croxson, M. S. 1974. Effect of single dose dexamethasone on the concentration of serum triiodothyronine in man. *J. Clin. Endocrinol. Metab.* 39:1151–54
59. Chopra, I. J., Williams, D. E., Orgiazzi, J., Solomon, D. 1975. Opposite effects of dexamethasone on serum concentrations of 3,3',5'-triiodothyronine (reverse T_3) and 3,3',5-triiodothyronine (T_3). *J. Clin. Endocrinol. Metab.* 41:911–20
60. Hershman, J. M., Pittman, J. A. 1971. Control of thyrotropin secretion in man. *N. Engl. J. Med.* 285:997–1005
61. Croxson, M. S., Hall, T. D., Nicoloff, J. T. 1976. Central inhibition of thyroid function by fasting. Program Ann. Meet. Endocr. Soc., 58th, p. 98 (Abstr.)
62. Bray, G. A. 1969. Effect of diet and triiodothyronine on the activity of sn-glycerol-3-phosphate dehydrogenase and on the metabolism of glucose and pyruvate by adipose tissue of obese patients. *J. Clin. Invest.* 48:1413–22

Ann. Rev. Med. 1977. 28:67–73
Copyright © 1977 by Annual Reviews Inc. All rights reserved

NEW CONCEPTS IN THE DIAGNOSIS OF PSYCHIATRIC DISORDERS

❖7223

Eli Robins, M.D.
Department of Psychiatry, Washington University School of Medicine,
St. Louis, Missouri 63110

INTRODUCTION

This paper presents new concepts in psychiatric diagnosis and does not limit itself to methods. These new concepts may be summarized as the use of a technique employing a categorical (typological) approach to psychiatric diagnosis, relating these concepts to the diagnosis of nonpsychiatric (medical) illness, describing the steps necessary and their consequences when these concepts are used, validation of psychiatric diagnosis, application of a few suggested means of other psychiatric classifications, and a discussion of heterogeneity and homogeneity in psychiatric classification and diagnosis.

CATEGORICAL (TYPOLOGICAL) DIAGNOSIS

Typological diagnosis in its present sense probably originated with Sydenham 300 years ago. The extension in this concept successfully spread to psychiatry in the nineteenth century, a little more than a century ago (Esquirol, Griesinger), but there is only a 75-year history of clear success in psychiatry (Kraepelin), in which the diagnoses are used as they are currently.

Steps in Psychiatric Diagnosis and Validation (1, 2)

Psychiatric diagnosis relies on a careful and comprehensive description of the clinical picture of the disorder. This may be a combination of clinical features thought to be associated with one another, or of a single striking clinical manifestation that has other clinical manifestations regularly associated with it. Demographic variables (race, age, and sex), age of onset of the illness (or symptoms), precipitating factors, and other parameters may also be used to define the clinical picture more precisely. The clinical picture thus does not include only symptoms.

Laboratory studies play only a small role in psychiatric diagnosis; the fully validated ones are limited chiefly to mental retardation and to organic brain syndromes. There is currently active work in a number of laboratries attempting to describe valid and reproducible laboratory results concerning the functional psychiatric disorders (3–23). So far there has not been an unchallenged validated finding concerning catecholamines, dopamine, acetylcholine, or serotonin levels, or metabolism in a functional psychiatric illness, e.g. schizophrenia or affective disorder.

A third consequence, frequently considered relatively unimportant, is the differentiation from other disorders of the disorder whose validation is being attempted. Since similar clinical pictures may be seen in patients suffering from different disorders (for example, cough and bloody sputum in lobar pneumonia, bronchiectasis, and bronchogenic carcinoma; or hallucinations in schizophrenia, manic-depressive illness, and organic brain syndrome), it is necessary to specify exclusion criteria so that patients with similar illnesses are not included in the group to be validated. These exclusion criteria should also permit exclusion of borderline patients and doubtful patients—i.e. an undiagnosed group—so that the index group for validation is as homogeneous as possible. The importance of having delimiting (exclusion) criteria should not be underemphasized. Without these delimiting criteria it is possible to have a mixed group—vitiating the validating effort, especially when the next two consequences are considered.

The fourth consequence, requiring the same care in interviews, is the follow-up study. The purpose of this part of the validation is to determine whether or not the original index patients are suffering from some other defined disorder that could account for the original clinical picture. If they are suffering from another such illness, the original patients probably did not constitute a homogeneous group, and it may be necessary to modify the diagnostic criteria. In the absence of a known cause or pathogenesis, as is true in the common psychiatric disorders, marked differences in outcome—between complete recovery and chronic illness, for example—suggest that the group is not homogeneous. The latter point is not so compelling, however, in suggesting diagnostic heterogeneity as is the finding of a change in diagnosis. The same illness may have a variable prognosis, but until we know more about the fundamental nature of the common psychiatric illnesses, marked differences in outcome should be regarded as a challenge to the validity of the original diagnosis. Thus the importance of step 3 is apparent, the exclusion of patients with illnesses in some ways similar to, but different from those in the index group.

The fifth consequence is a family study, requiring as careful interviews as the index interview, and requiring more interviews because there are more family members to be interviewed. Most psychiatric illnesses have been shown to run in families, whether the investigations were designed to study hereditary or environmental causes. Independent of etiology, therefore, the finding of an increased incidence of the same disorder among the close relatives of the original patients strongly indicates that one is dealing with a valid entity. These family interviews are family studies in which the surviving relatives are interviewed personally. These are not family-history studies in which a family history is obtained from, for example, the index patient and one other relative about the other family members. They require a personal interview with each surviving family member, an examination of the death certificate of deceased members, and an interview with the last physicians of index patients who have died.

APPLICATION OF TYPOLOGICAL CONCEPTS TO DIAGNOSIS OF NONPSYCHIATRIC ILLNESS

The five steps in diagnosis and validation, described in a somewhat different context, have been used in the study of nonpsychiatric medical diagnoses (24–28).

DEFINITION OF PSYCHIATRIC ILLNESSES AND SYMPTOMS COMPRISING THEM

Definition of Psychiatric Illness

Definitions of a number of psychiatric illnesses have been detailed (29). These definitions consist of a set of criteria for each illness defined. An example of one such illness defined categorically (typologically) is primary affective disorder (depressive). The definition consists of specified criteria, most of which are symptoms and some of which may include demographic and chronological definitions, as follows: For a diagnosis of depression, criteria 1 through 3 are required.

1. Dysphoric mood characterized by emotional symptoms such as the following: depressed, sad, blue, despondent, hopeless, "down in the dumps," irritable, fearful, worried, or discouraged.
2. At least five of the following criteria are required for "definite" depression; four are required for "probable" depression: (*a*) poor appetite or weight loss (positive if two pounds a week or ten pounds or more a year when not dieting); (*b*) sleep difficulty (include insomnia or hypersomnia); (*c*) loss of energy, e.g. fatigability, tiredness; (*d*) agitation or retardation; (*e*) loss of interest in usual activities, or decrease in sexual drive; (*f*) feelings of self-reproach or guilt (either may be delusional); (*g*) complaints of, or actually diminished ability to think or concentrate, such as slow thinking or mixed-up thoughts; and (*h*) recurrent thoughts of death or suicide, including thoughts of wishing to be dead.
3. A psychiatric illness lasting at least one month with no preexisting psychiatric conditions such as schizophrenia, anxiety neurosis, phobic neurosis, obsessive-compulsive neurosis, hysteria, alcoholism, drug dependency, antisocial personality, homosexuality and other sexual deviations, mental retardation, or organic brain syndrome. (Patients with life-threatening or incapacitating medical illness preceding and paralleling the depression do not receive the diagnosis of primary depression.) The information is elicited by interviewing and examining the patient and at least one relative or friend who knows the patient well.

Definition of a Symptom

To make defined changes, or to retain already adequately defined illnesses, requires that the symptoms clustering to define an illness also be defined. These symptoms are currently most simply defined by eliciting the symptom or behavior from the patient (or his relative) and then deciding if the symptom is to be scored as present or absent. A symptom that a patient says he has is still scored as absent if it does not occur more frequently or severely than in the patient's peers. As an example, a positive answer to the question, "Do you suffer from headaches?", is scored as absent if the headaches do not differ in frequency or severity from those suffered by most of the patient's peers. A single mild headache occurring quarterly or annually, and not a new symptom, is not sufficiently different from that found in most of the population to deserve a positive score, even though the patient may have such mild headaches. Since this decision about scoring relies on clinical judgment of varying subtlety, depending on which symptoms and which characteristics manifest themselves, less subtle scoring techniques are needed. These techniques have been outlined (2, 29, 30). A symptom is defined as being present (or absent) if it meets the five following criteria, and the two instructions for recording a symptom have been followed.

1. The symptom led the patient to go to a physician (or osteopath or other professional) on more than one occasion.
2. The symptom was disabling. Disability is defined as the patient's report that the symptom caused him to make changes (even minor ones) in his life because of the symptom. These changes include job troubles, trouble sleeping, need for rest, need for help when such was not ordinarily needed, restriction of social and recreational life, and a belief by the patient that the symptom definitely made a difference in his life even though he cannot pinpoint the difference. The symptom, because it was disabling, necessitated the patient to make these changes. They are not primary problems themselves, but brought about by the disability caused by the symptoms.
3. The symptom led the patient to take medication (excluding aspirin and closely related drugs, and laxatives) on more than one occasion and did so more often than semiannually.
4. The physician believes that because of its clinical importance, the symptom should be scored as positive, even though it does not fulfill criteria 1, 2, or 3. For example, a spell of blindness lasting a few minutes, which the patient may minimize—i.e. deny that criteria 1 through 3 apply—would be scored as positive. This standard is also meant to include any symptom that the physician believes to be of clinical significance to him, without regard to its fulfilling criteria 1, 2, and 3.
5. A symptom is not scored as positive if it is explicable by known nonpsychiatric disease of the patient. The instructions for recording responses to questions about a symptom are (*a*) any positive response must be elaborated with regard to its qualitative description, its frequency, and its chronology; (*b*) any item to which a patient gives a putatively positive response, but which is scored as negative, must be elaborated so that a research group or a group of clinicians can reach its own decision concerning the validity of the negative scoring.

ASSESSMENT OF HOMOGENEITY OR HETEROGENEITY OF THE DIAGNOSES MADE

Homogeneity is determined by clinical picture, follow-up study, and family study. The clinical picture in a homogeneous group should closely correspond among patients with a given diagnosis. Follow-up study should indicate there has not been a change to another diagnosis (other than becoming well) since the time of initial examination [See the fourth step in psychiatric diagnosis and validation (p. 68).] If there is not a change in diagnosis, that is the most compelling follow-up finding indicating diagnostic homogeneity. If there has been a change in diagnosis, that finding strongly indicates diagnostic heterogeneity.

The family studies of a series of patients with the same initial diagnosis should indicate an increased incidence of the illness in the relatives of the index case. This finding strongly indicates diagnostic homogeneity of the original group. The failure to find an increased family incidence or prevalence in the families of the index group equally strongly indicates diagnostic heterogeneity. (See step 5, p. 68.) The laboratory studies in which diagnostic heterogeneity has been inferred (7, 13, 20, 21, 27, 28) are not presently a sufficiently good reason for assuming that the findings support diagnostic heterogeneity or homogeneity, because the laboratory study of functional psychiatric illness is still in a primitive state. (See step 2, p. 68.) Decisions regarding heterogeneity or homogeneity of an index sample should not be based on these still-unvalidated findings, especially when the follow-up results and the family study both indicate either homogeneity or heterogeneity of the original sample.

The importance of validating the homogeneity or heterogeneity of patients comprising the original sample is that a heterogeneous group limits the valid generalizations that may be made from the original sample of patients and vitiates these generalizations. In contrast, a demonstrated homogeneous original group makes generalizations valid and permits accurate and meaningful generalizations.

CONCLUSION

The salient points in this paper include the following: (*a*) a description of categorical (typological) diagnosis, applicable to psychiatric as well as to medical illness; (*b*) description and discussion of the five steps necessary to validate clinically a categorical (typological) diagnosis; (*c*) a description of the currently primitive state of laboratory studies of functional disorders (e.g. schizophrenia, affective disorders, and other functional disorders); (*d*) a definition of various psychiatric illnesses and a description of the criteria necessary to qualify a symptom for describing them; and (*e*) the means and purposes of assessing diagnostic homogeneity or heterogeneity in psychiatric illnesses.

ACKNOWLEDGMENTS

These studies were supported in part by grants from the National Institute of Mental Health MH 13002 and MH 25430. The latter grant is part of the NIMH Collaborative Study on the Psychobiology of Depression: Clinical Pilot Study.

Literature Cited

1. Robins, E., Guze, S. B. 1970. Establishment of diagnostic validity in psychiatric illness: Its application to schizophrenia. *Am. J. Psychiatry* 126:983–88
2. Robins, E., Gentry, K., Munoz, R. A., Marten, S. 1976. Study and 18-month follow-up of 314 psychiatric emergency room patients: A contrast of the three more common illnesses with the 10 less common. III. Findings of the follow-up. *Arch. Gen. Psychiatry.* In press
3. Greenspan, K., Schildkraut, J. J., Gordon, E. K., Baer, L., Aronoff, M.S., Durell, J. 1970. Catecholamine metabolism in effective disorders: III. 3-methoxy-4-hydroxyphenylglycol and other catecholamine metabolites in patients treated with lithium carbonate. *J. Psychiatr. Res.* 7:171–83
4. Bond, P. A., Jenner, F. A., Sampson, G. A. 1972. Daily variations of the urine content of MHPG in two manic-depressive patients. *Psychol. Med.* 2:81–85
5. Jones, F. D., Maas, J. W., Dekirmenjian, H., Fawcett, J. A. 1973. Urinary catecholamine metabolites during behavioral changes in a patient with manic-depressive cycles. *Science* 179:300–2
6. Shopsin, B., Wilk, S., Gershon, S., Roffman, M., Goldstein, M. 1973. In *Frontiers in Catecholamine Research.* ed. E. Usdin, S. Snyder, pp. 1173–79. New York: Pergamon. 1229 pp.
7. Shaw, D. M., O'Keeffe, R., MacSweeney, D. A., Brooksbank, B. W. L., Noguera, R., Coppen, A. 1973. 3-methoxy-4-hydroxyphenylglycol in depression. *Psychol. Med.* 3:333–36
8. Dekirmenjian, H., Maas, J. W., Fawcett, J. 1973. Urinary excretion of norepinephrine and its metabolites in human control subjects. Presented at Ann. Meet. Am. Psychiatr. Assoc., Honolulu, Hawaii
9. Maas, J. W., Dekirmenjian, H., Deleon-Jones, F. 1973. See Ref. 6, pp. 1091–96
10. Wise, C. D., Stein, L. 1973. Dopamine-β-hydroxylase deficits in the brains of schizophrenic patients. *Science* 181: 344–47
11. Ebert, M. H., Post, R. M., Goodwin, F. K. 1972. Effect of physical activity on urinary MHPG excretion in depressed patients. *Lancet* 2:766
12. Olson, L., Nystrom, B., Seiger, A. 1973. See Ref. 6, pp. 1097–1100
13. Shopsin, B., Wilk, S., Gershon, S., Davis, K., Suhl, M. 1973. Cerebrospinal fluid MHPG: An assessment of norepinephrine metabolism in affective disorders. *Arch. Gen. Psychiatry* 28:230–33
14. Maas, J. W., Fawcett, J. A., Dekirmenjian, H. 1972. Catecholamine metabolism, depressive illness and drug response. *Arch. Gen. Psychiatry* 26: 252–62
15. Schildkraut, J. J. 1973. Norepinephrine metabolites as biochemical criteria for classifying depressive disorders and predicting responses to treatment: Preliminary findings. *Am. J. Psychiatry* 130:695–98
16. Beckmann, H., Goodwin, F. K. 1975. Antidepressant response to tricyclics and urinary MHPG in unipolar patients. *Arch. Gen Psychiatry* 32:17–21
17. Janowsky, D. S., Davis, J. M., El-Yousef, M. K., Sekerke, H. J. 1972. A cholinergic-adrenergic hypothesis of mania and depression. *Lancet* 2:632–35
18. van Praag, H. M., Korf, J. 1974. Serotonin metabolism in depression: Clinical application of the probenecid test. *Int. Pharmacopsychiatry* 9:35–51
19. Asberg, M., Bertilsson, L., Tuck, D., Cronholm, B., Sjoquist, F. 1973. Indoleamine metabolites in the cerebrospinal fluid of depressed patients before and during treatment with nortriptyline. *Clin. Pharmacol. Ther.* 14:277–86
20. Maas, J. W. 1975. Biogenic amines and depression: Biochemical and pharmacological separation of two types of depression. *Arch. Gen. Psychiatry* 32: 1357–61
21. Asberg, M., Thoren, P., Traskman, L. 1976. "Serotonin depression"—A biochemical subgroup within the affective disorders? *Science* 191:478–80
22. Grote, S. S., Moses, S. G., Robins, E., Hudgens, R. W., Croninger, A. B. 1974. A study of selected catecholamine metabolizing enzymes: A comparison of depressive suicides and alcoholic suicides with controls. *J. Neurochem.* 23:791–802
23. Moses, S. G., Robins, E. 1975. Regional distribution of norepinephrine and dopamine in brains of depressive suicides and alcoholic suicides. *Psychopharmacol. Comm.* 1:327–37
24. Ressler, N. 1975. Computer-assisted diagnosis by a model-free system of direct data analysis. *Perspect. Biol. Med.* 19:101–17
25. Heller, P., Winstein, H. G., West, M., Zimmerman, H. J. 1960. Enzymes in anemia: A study of abnormalities of sev-

eral enzymes of carbohydrate metabolism in the plasma and erythrocytes in patients with anemia, with preliminary observations of bone marrow enzymes. *Ann. Intern. Med.* 53:898–913.

26. Tonks, D. B. 1963. A study of the accuracy and precision of clinical chemistry determinations in 170 Canadian laboratories. *Clin. Chem.* 9:217–30
27. Preston, J. A., Batsakis, J. G., Briere, R. O., Taylor, R. V. 1965. Serum creatine phosphokinase. A clinical and laboratory evaluation. *Am. J. Clin. Pathol.* 44:71–76
28. Norris, A. H., Shock, N. W. 1966. Aging and variability. *Ann. NY Acad. Sci.* 134:591-601
29. Feighner, J. P., Robins, E., Guze, S. B., Woodruff, R. A. Jr., Winokur, G., Munoz, R. 1972. Diagnostic criteria for use in psychiatric research. *Arch. Gen. Psychiatry* 26:57–63
30. Robins, E. 1976. Categories versus dimensions in psychiatric classification. *Psychiatr. Ann.* 6:368–74

Ann. Rev. Med. 1977. 28:75–83
Copyright © 1977 by Annual Reviews Inc. All rights reserved

EFFECT OF ANESTHETIC DRUGS ON MYOCARDIAL PERFORMANCE IN MAN[1]

◆7224

Robert G. Merin, M.D.
Departments of Anesthesiology and Pharmacology, and Toxicology, University of Rochester School of Medicine and Dentistry, Rochester, New York, 14642

Although drugs that depress the central nervous system (CNS) sufficiently to produce surgical anesthesia also depress cardiac muscle in vitro, direct transposition from muscle strips to man is not possible for a variety of reasons (1). Consequently, in order to study the effects of anesthetics in man, we must be able to quantitate ventricular function in man.

Both isolated muscle and the intact heart function in two ways. When the muscle fiber is not allowed to shorten, but only develops tension, muscle contracts isometrically (Figure 1). The corollary in the intact heart is ventricular pressure development before the aortic (or pulmonary) valve opens. In man this aspect of ventricular function has been estimated by the tension developed between the "feet" of a strain-gauge arch (SGA) or measurement of pressure and time indices before the ventricular outflow valves open (2). During isotonic contraction, the muscle shortens, moving a load (afterload) and performing work (Figure 1). The analogy in the intact heart is the ejection of blood from the ventricles. This "pump" function can be estimated in man by measuring cardiac output (CO), the time and volume ejection indices, the I-J wave of the ballistocardiogram, or aortic pressure pulse analysis (3). Both types of function are dependent on several factors other than the intrinsic function of the contractile element in the muscle:

1. An increase in the resting (diastolic) muscle length or *preload* (Figure 1, top) leads to increased function of both isolated muscle and the intact heart.
2. As *afterload* increases, the speed that the isolated muscle can move is decreased. Likewise, in the intact heart, the higher the afterload (force against which the ventricle must eject blood), the lower the cardiac output, if other influences do

---

[1] Supported in part by USPHS grant #HL-31752.

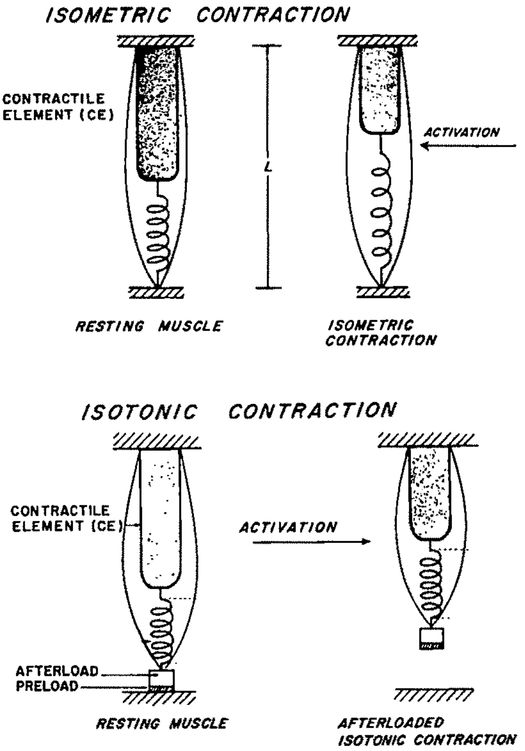


Figure 1. Schema of muscle function: (*a*) isometric—resting muscle length (1) unchanged, tension developed; (*b*) isotonic—resting muscle length shortens, load (after) is moved. Adapted from (1) with permission.

not change. However, isometric or "muscle" function in the intact heart may respond differently to afterload changes. If afterload increases, more pressure must be developed in the ventricle in order to open the outflow valves. Consequently, the measures of muscle function in the intact heart may increase as afterload increases.

3. Increase in the rate of stimulation will increase performance in isolated muscle. Sometimes in the intact heart, heart-rate increases lead to decreased diastolic filling (decreased preload) and the positive inotropic effect may be negated.
4. Of course, the bathing fluid (coronary blood supply for the intact heart) is important for the level of function. Table 1 indicates how these influences can be measured in man, in decreasing order of precision for each. Although it is practically impossible to control these important factors in man, they should at least be measured as accurately as possible. Otherwise, reliable estimates of the effect of drugs on cardiac function are not possible. Unfortunately, there are few studies of the effect of anesthetics on man where all of these influences have been measured, much less controlled.

Table 1 Factors influencing cardiac function

Isolated muscle	Man (in order of decreasing precision)
Preload	ventricular end diastolic volume ventricular end diastolic pressure atrial pressure venous pressure
Afterload	aortic (pulmonary–artery) diastolic pressure (isometric) systemic (pulmonary) vascular resistance (impedance) arterial pressure
Rate of stimulation	heart rate
Bathing fluid	coronary blood

INHALATION ANESTHETICS

Several potent inhalation anesthetics produce minimal evidence of cardiac depression in man unless gross overdoses are given. Skovsted & Price observed that diethyl ether, cyclopropane, and fluroxene produce evidence of central sympathetic excitation, and classified them as "Group I" anesthetics. Other inhaled anesthetics, including halothane, methoxyflurane, and enflurane failed to produce this effect but produce dose-related cardiac depression in man and have been classified as "Group II" anesthetics (4). In addition, the Group I drugs are associated with a high incidence of some unpleasant side effects such as nausea and delirium. They are all flammable, so that their clinical usefulness is limited at the present time.

Diethyl Ether

As with most of the anesthetics, only "pump" function has been measured with diethyl ether and little documentation of preload has been made. CO has been little affected by high diethyl ether concentrations, mostly as a result of increased heart rate (5, 6). In the best published study, ventricular ejection rate was unchanged (5). Markedly flattened ventricular function curves produced by high ether concentrations in the dog could be restored to control by mechanically elevating aortic pressure (7). The observation of elevated catecholamine concentrations in the plasma (5) and the marked reduction in ventricular function produced by β-adrenergic blockade during ether anesthesia in man (8) suggests that sympathetic nervous stimulation is an important mechanism in the maintenance of cardiac homeostasis during ether anesthesia.

Cyclopropane

Again, only pump function has been reported with cyclopropane. Preload has been poorly documented (right atrial pressure), but was consistently elevated. Heart rate was generally unchanged and afterload was also increased. Under these conditions there was no significant change in CO or stroke volume (5). Local anesthetic blockade of cardiac sympathetic ganglia depressed stroke volume at the same or increased right atrial pressures during cyclopropane anesthesia in man (9). Elevated plasma catecholamine concentrations have been found (5). Consequently, it would seem that Price's hypothesis is tenable with cyclopropane as well as with diethyl ether. However, conflicting reports in animals have left some doubt as to the mechanism of the effect of cyclopropane on cardiac function in man (10, 11).

Fluroxene

The only halogenated anesthetic in Group I, fluroxene, maintained cardiac output unchanged at very high doses in human volunteers (5). Both heart rate and plasma norepinephrine concentrations were markedly increased, although no other studies have been reported to support these observations. Although fluroxene is considerably less flammable than diethyl ether or cyclopropane, it has never been widely used and commercial production has currently been halted.

Halothane

The most popular of the inhalation anesthetics, halothane, has also been the most widely studied. Whether observing muscle function with a SGA (12) or right ventricular dP/dt (13); or measuring pump function with CO (5,13), ejection indices (5), or the BCG (5), all studies have shown a dose-related decrease in cardiac performance with halothane. More precise animal studies have further documented this effect (14,15). Mechanical elevation of decreased aortic pressure during high-dose halothane anesthesia in dogs did not increase ventricular function, in contrast to the results mentioned above with diethyl ether (7,15).

Methoxyflurane

Although less well studied, the effects of this potent, slowly acting, long-lasting halogenated ether appear to resemble those of halothane (16).

Enflurane and Isoflurane

The published studies on the cardiovascular effects of enflurane in man are all open to criticism (17, 18). However, a carefully done experiment with human volunteers has indicated that the drug produces cardiac depression equivalent to that seen with halothane (N.T. Smith et al, personal communication). Animal work supports this conclusion (19). Isoflurane, an investigational isomer of enflurane, appears to produce significantly less cardiac depression in man (20) and in the dog (R.G. Merin, unpublished observations). The mechanism behind the difference in the two drugs remains unclear at present.

Nitrous oxide

The majority of anesthetics administered in the Western world today include nitrous oxide as an adjuvant, primarily because the gas has been thought to have no pharmacologic effects other than amnesia and analgesia. However, as little as 40% nitrous oxide alone has been shown to depress ventricular function (CO and BCG measurements) in normal volunteers (21). When combined with other more potent drugs, myocardial depression (morphine) (22) or apparent peripheral vasocontriction (halothane) (23) may be produced. The mechanisms behind these diverse effects are probably central in origin, but have not been clarified as yet. It seems that the old opinion about the benign pharmacologic nature of this most widely used of the anesthetic gases must be modified in present practice.

INTRAVENOUS DRUGS

None of the intravenous drugs available for the clinician in the United States today can produce all the requirements of surgical anesthesia. The hypnotics and tranquilizers lack analgesic and muscle-relaxing properties. The analgesics and neuromuscular blocking drugs do just what their name implies. Ketamine comes the closest to fulfilling the criteria, but if muscle relaxation is necessary this drug must also be supplemented. Consequently, the intravenous drugs are almost always used in combination and/or with inhalation or regional anesthesia.

Hypnotics

These drugs are primarily used for induction of anesthesia. The ultra–short-acting thio- and oxybarbiturates (thiopental, thiamylal, methohexital) remain the standard drugs. Early work suggested that these drugs produced venodilation predominantly and decreased CO through a decrease in venous return (24). Although more recent studies with systolic time intervals (25) and left ventricular (LV) pressure derivatives (26,27) have suggested a direct negative inotropic effect with clinical doses in man, preload was not always documented. In the best studies, from Göttingen, West Germany, significant tachycardia probably decreased the observed negative inotropic effect (26,27). The same group has seen similar effects with several drugs not available in the United States, including propranidid (a eugenol), and althesin (a corticosteroid combination). Of all induction agents studied by the Göttingen group, only the experimental imidazoyl-carboxyl ester, etomidate, failed to produce significant tachycardia and some evidence of ventricular depression. Diazepam has become very popular both for induction and supplementation of other anesthetic techniques. Low doses (0.1 mg/kg) (28) produced small decreases in preload, afterload, and stroke volume. Large doses (0.77 mg/kg) (29) resulted in more striking decreases in arterial pressure and stroke volume, now accompanied by an increase in heart rate. Diazepam is doubtless capable of causing significant depression of cardiac performance, as are all the available drugs of this type.

Analgesics

Although small doses of narcotic analgesics such as morphine and meperidine have been used for decades, only since the studies of Lowenstein et al in 1969 (30) has an attempt been made to use this class of drugs as the primary anesthetic. In that study, 1.0–2.0 mg/kg of morphine produced little change in pump function in healthy patients as measured by CO derivatives. More importantly, 1.0 mg/kg actually increased CO in patients with valvular heart disease (VHD). Another study in healthy volunteers supported the original data (22), although in clinical practice tachycardia and both hypo- and hypertension have been occasional problems, especially in patients with ischemic heart disease (IHD). Although the new synthetic narcotic analgesic fentanyl is being widely used, especially in patients with heart disease, there is no objective evidence that the cardiac effects differ significantly from those reported with morphine (31,32).

Ketamine

One of the original attractions of ketamine was the lack of a negative inotropic effect. In fact, ketamine appeared to increase cardiac performance in healthy patients (33). The best of the studies published on the effect of anesthetics on ventricular function in man investigated the effect of 2 mg/kg of ketamine in angina pectoris patients without demonstrated coronary pathology on angiogram (34). Tweed et al showed a definite increase in "muscle" function (using Mason's Common Peak Intraventricular Pressure technique) (3) and cardiac output, while keeping heart rate constant by atrial pacing (Figure 2).

Neurolept Anesthesia

The combination of potent tranquilizers such as phenothiazines and butyrophenones with potent analgesics has been termed neuroleptic anesthesia. Although the most popular combination, droperidol and fentanyl, produced little effect on ventricular function in healthy man, the studies were all without preload measurement (26,27,31,32). A definitive statement must await further study.

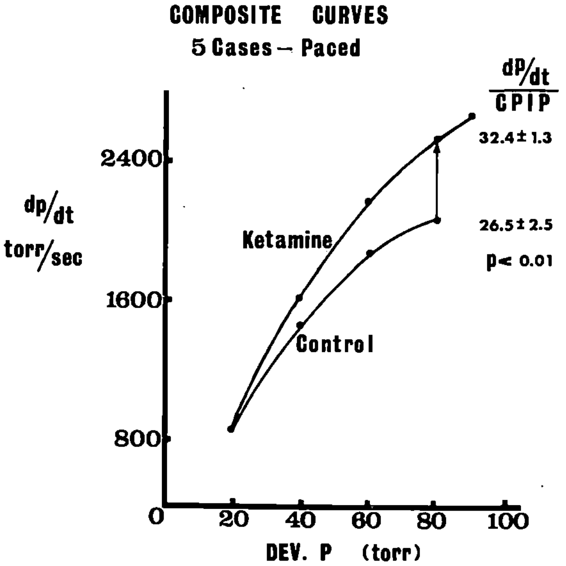


Figure 2. Effect of ketamine 2 mg/kg on cardiac muscle function in man at constant heart rate. DEV.P=developed pressure in the left ventricle; dp/dt=rate or rise of left ventricular pressure; and CPIP = Common Peak Intraventricular Pressure (used with permission of W.A. Tweed, M.D).

Neuromuscular Blocking Drugs

The ability to safely anesthetize and operate on severely ill patients is due in large measure to the development of this class of drugs. No longer are high doses of potent CNS (and cardiovascular) depressants necessary to provide an adequate surgical field. Obviously, then, these drugs must have minimal negative inotropic effects. Using a SGA on the right ventricles of patients during cardiopulmonary bypass, Longnecker et al demonstrated that this is indeed true for D-tubocurarine (35) and gallamine (36). Studies in intact man have supported these observations (37). Gallamine and pancuronium frequently increase heart rate, and succinylcholine and D-tubocurarine can produce a bradycardia, probably through effects on the autonomic nervous system (ANS). In a variety of disease states, succinylcholine is also capable of eliciting enough potassium release from damaged muscle to severely depress ventricular function (38).

REGIONAL ANESTHESIA

There are two ways in which the use of local anesthetics to block nerve transmission can affect the heart. If large doses of the drug are used (or if the drugs are injected directly intravascularly), the direct effect of the local anesthetic on the heart may depress ventricular performance (39) (although usually CNS toxicity will appear first) (40). If the nerve block includes a significant portion of the ANS (spinal, epidural, and IX, X cranial nerves), cardiac performance may also be affected. However, in healthy, normovolemic man, epidural and spinal anesthesia to as high as the second thoracic dermatome produced minimal changes in the pump function of the heart (41, 42).

SUMMARY

All potent CNS depressant drugs can depress cardiac function in man in a dose-dependent manner. The dose-effect curve is considerably flatter with several drugs (diethyl ether, cyclopropane, fluroxene, isoflurane, and ketamine), presumably from sympathetic nervous-system activation. Potent analgesics and tranquilizers appear to produce less depression, but have been incompletely studied. Neuromuscular blocking drugs and regional anesthesia produce minimal effects on the heart in healthy people. However, not as much is known about diseased man. For instance, nitrous oxide produces more depression in "muscle" function in IHD patients (43), while diazepam (28) and morphine (44) do not adversely affect pump function in this class of patients. Fluroxene (45) is more depressant in VHD patients, but nitrous oxide (46), morphine (30), fentanyl (46), and droperidol-fentanyl (46) seem to have equivalent effects to those seen in healthy patients. In any given patient, therefore, accurate prediction of the effect of any anesthetic drug on cardiac performance is not possible. Adequate monitoring and careful titration of drug dose offer the safest method of assuring a satisfactory response.

Literature Cited

1. Shimosato, S., Etsten, B. E. 1969. Effects of anesthetic drugs on the heart: A critical review of myocardial contractility and its relationship to hemodynamics. *Clin. Anesth.* 3:17–72
2. Braunwald, E., Ross, J., Gault, J. H., Mason, D. T., Mills, C., Gabe, I. T., Epstein, S. E. 1969. Assessment of cardiac function. *Ann. Intern. Med.* 70:369–99
3. Merin, R. G. Physiological functions during anesthesia. Cardiac physiology and cardiac output. In *Anesthesiology,* ed. C. P. Larson, H. Wollman. Philadelphia: Lippincott. In press
4. Skovsted, P., Price, H. L. 1972. Effects of Ethrane on arterial pressure, preganglionic sympathetic activity and barostatic reflexes. *Anesthesiology* 36:257–62
5. Eger, E. I., Smith, N. T., Cullen, D. J., Cullen, B. F., Gregory, G. A. 1971. Comparison of the cardiovascular effects of halothane, fluroxene, ether and cyclopropane in man. *Anesthesiology* 34:25–41
6. Kubota, Y., Schweizer, H. J., VanDam, L. D. 1962. Hemodynamic effects of diethyl ether in man. *Anesthesiology* 23:306–11
7. Etsten, B. E., Shimosato, S. 1966. Influence of stress on the performance of the heart during ether and halothane anesthesia. *Acta Anaesthesiol. Scand.* (Proc. I, Suppl.) 23:242–45
8. Jorfeldt, L., Löfstrom, B., Möller, J., Rosén, A. 1967. Cardiovascular pharmacodynamics of propranolol during ether anaesthesia in man. *Acta Anaesthesiol. Scand.* 11:159–65
9. Price, H. L., Jones, R. E., Deutsch, S., Linde, H. W. 1962. Ventricular function and autonomic nervous activity during cyclopropane anesthesia in man. *J. Clin. Invest.* 41:604–10
10. Markee, S., Wang, H. H., Wang, S. C. 1966. Effects of cyclopropane on the central vasomotor mechanism of the dog. *Anesthesiology* 27:742–55
11. Fukunaga, A., Epstein, R. M. 1974. Effects of cyclopropane on the sympathetic nervous system and on neural regulation of the circulation in the cat. *Anesthesiology* 40:323–35
12. Mahaffey, J. C., Aldinger, E. E., Sprouse, J. H., Darby, T. D., Thrower, W. B. 1961. Cardiovascular effects of halothane. *Anesthesiology* 22:982–86
13. Price, H. L., Skovsted, P., Pauca, A. L., Cooperman, L. H. 1970. Evidence for beta-receptor activation produced by halothane in normal man. *Anesthesiology* 32:389–95
14. Vatner, S. F., Smith, N. T. 1974. Effect of halothane on left ventricular function and distribution of regional blood flow in dogs and primates. *Circ. Res.* 34:155–67
15. Merin, R. G., Kumazawa, T., Luka, N. L. 1976. Myocardial metabolism and function in the conscious dog and during halothane anesthesia. *Anesthesiology* 44:402–15
16. Libonati, M., Cooperman, L. H., Price, H. L. 1971. Time dependent circulatory effects of methoxyflurane in man. *Anesthesiology* 34:439–46
17. Graves, C. L., Downs, N. H. 1974. Cardiovascular and renal effects of enflurane in surgical patients. *Anesth. Analg. Cleveland* 53:898–903
18. Levesque, F. R., Naragas, V., Shanks, C., Shumosato, S. 1974. Circulatory effects of Ethrane in human volunteers at constant carbon dioxide tension. *Can. Anaesth. Soc. J.* 21:580–85
19. Merin, R. G., Kumazawa, T., Luka, N. L. 1976. Enflurane depresses myocardial function, perfusion and metabolism in the dog. *Anesthesiology.* 45:501–7
20. Stevens, W. C., Cromwell, T. H., Halsey, M. J., Eger, E. I., Shakespeare, T. F., Bahlman, S. H. 1971. Cardiovascular effects of a new inhalation anesthetic, Forane, in human volunteers at constant arterial carbon dioxide tension. *Anesthesiology* 35:8–16
21. Eisele, J. H., Smith, N. T. 1972. Cardiovascular effects of 40% nitrous oxide in man. *Anesth. Analg. Cleveland* 51: 956–61
22. Wong, K. C., Martin, W. E., Hornbein, T. F., Freund, F., Everett, J. 1973. Cardiovascular effects of morphine sulphate with oxygen and nitrous oxide in man. *Anesthesiology* 38:542–49
23. Smith, N. T., Eger, E. I., Stoelting, R. K., Whayne, T. F., Cullen, D., Kadis, L. B. 1970. Cardiovascular and sympathomimetic responses to the addition of nitrous oxide to halothane in man. *Anesthesiology* 32:410–21
24. Etsten, B. E., Li, T. H. 1955. Hemodynamic changes during thiopental anesthesia in humans. *J. Clin. Invest.* 34:500–19

25. List, W. F., Hiotakis, K., Gravenstein, J. S. 1972. Die Werkung von Thiopental auf Myokardfunktion. *Anaesthetist* 21:388–90
26. Sonntag, H. 1973. Koranardurchblutung und Energieumsatz des menschlichen Herzen unter verschiedenen Anaesthetika. *Anaesthesiol. Wiederbeleb.* 79 pp.
27. Kettler, D., Sonntag, H. 1974. Intravenous anesthetics: Coronary, blood flow and myocardial oxygen consumption. *Acta Anaesthesiol. Belg.* 25:384–401
28. Cote, P., Gueret, P., Bourassa, M. G. 1974. Systemic and coronary hemodynamic effects of diazepam in patients with normal and diseased coronary arteries. *Circulation* 50:1210–16
29. Rao, S., Sherbaniuk, R. W., Prasad, K., Lee, S. J. K., Sproule, B. J. 1973. Cardiopulmonary effects of diazepam. *Clin. Pharmacol.* 14:182–90
30. Lowenstein, E., Hollowell, P., Levine, F. H., Daggett, W. M., Austen, W. G., Laver, M. B. 1969. Cardiovascular response to large doses of intravenous morphine in man. *N. Engl. J. Med.* 281:1389–93
31. Ferrari, H. A., Bortan, R. J., Talton, I. H., Cannett, R., Goodrich, J. R. 1974. Action of droperidol and fentanyl on cardiac output and related hemodynamic variables. *South. Med. J.* 67: 49–53
32. Graves, C. L., Downs, N. H., Browne, A. B. 1975. Cardiovascular effects of minimal analgesic qualities of Innovar, fentanyl and droperidol in man. *Anesth. Analg. Cleveland* 54:15–23
33. Virtue, R. W., Alanis, J. M., Mori, M., Lafargue, R. T., Vogel, J. H. K., Metcalf, D. R. 1968. An anesthetic agent: 2-orthochlorophenyl, 2-metheglamino cyclohexanone HCl (CI-581) *Anesthesiology* 28:823–33
34. Tweed, W. A., Mymin, D. 1974. Myocardial force-velocity relationship during ketamine anesthesia at constant heart rate in man. *Anesthesiology* 41:59–64
35. Longnecker, D. E., Stoelting, R. K., Morrow, A. G. 1970. Cardiac and peripheral vascular effects of D-tubocurare in man. *Anesth. Analg. Cleveland* 49:660–64
36. Longnecker, D. E., Stoelting, R. K., Morrow, A. G. 1973. Cardiac and peripheral vascular effects of gallamine in man. *Anesth. Analg. Cleveland* 52: 931–35
37. Stoelting, R. K. 1972. Hemodynamic effects of pancuronium and D-tubocurare in anesthetized patients. *Anesthesiology* 36:612–15
38. Cooperman, L. H. 1970. Succinylcholine induced hyperkalemia in neuromuscular disease. *J. Am. Med. Assoc.* 213:1876–71
39. Binnion, P. F., Murtagh, G., Pollock, A. M., Fletcher, E. 1969. Relation between plasma lignocaine level and induced haemodynamic changes. *Br. Med. J.* 3:390–92
40. Jorfeldt, L., Lofstrom, B., Pernow, B., Persson, B., Wahren, J., Widman, B. 1968. Effect of local anaesthetics on the central circulation and respiration in man and dog. *Acta Anaesthesiol. Scand.* 12:153–69
41. Shimosato, S., Etsten, B. E. 1969. Role of the venous system in cardio-circulatory dynamics during spinal and epidural anesthesia in man. *Anesthesiology* 30:619–28
42. Bonica, J. J., Berges, P. U., Morikawa, K. 1970. Circulatory effects of peridural block: I. Effects of level of analgesia and dose of lidocaine. *Anesthesiology* 33:619–26
43. Eisele, J. H., Reitan, J. A., Massumi, R. A., Zelis, R. F., Miller, R. R. 1976. Myocardial performance and N_2O analgesia in coronary artery disease. *Anesthesiology* 44:16–20
44. Lappas, D. G., Geha, D., Fischer, J. E., Laver, M. B., Lowenstein, E. 1975. Filling pressures of the heart and pulmonary circulation of the patient with coronary artery disease after large intravenous doses of morphine. *Anesthesiology* 43:153–59
45. Wiberg-Jorgensen, F., Skovsted, P., Misfeldt, B. 1973. Cardiovascular hemodynamics during fluroxene anesthesia in patients with aortic valvular disease. *Acta Anaesthesiol. Scand.* 17:136–41
46. Stoelting, R. K., Gibbs, P. S., Creasser, C. W., Peterson, C. 1975. Hemodynamic and ventilatory responses to fentanyl, fentanyl-droperidol, and nitrous oxide in patients with acquired valvular heart disease. *Anesthesiology* 42:319–24

Ann. Rev. Med. 1977. 28:85–90
Copyright © 1977 by Annual Reviews Inc. All rights reserved

DIAGNOSIS AND TREATMENT OF FAT EMBOLISM

❖7225

Joseph A. Moylan, M.D.
Department of Surgery, Duke University Medical Center, Durham, North Carolina 27710

Merle A. Evenson, Ph.D.
Department of Medicine, University of Wisconsin Hospitals, Madison, Wisconsin 53706

Pulmonary insufficiency remains a major complication following multisystem trauma. The fat emboli syndrome is a leading cause of this respiratory problem occurring in the early postaccident period. Ever since Zenker (1) first described postmortem findings of fat emboli in the lungs of accident victims, there has been considerable interest by traumatologists in defining the pathophysiology, prevention, and treatment of the clinical entity. Because of its protean manifestation, the syndrome often goes unnoticed until the terminal phase.

Although experimental work with fat embolism dates back more than two centuries, it was not until the late 1800s that the initial clinical description of the triad of confusion, dyspnea, and petechiae was described in German literature by Von Bergmann (2). Soon after, the first American experience with this posttraumatic disease process was recorded by Fenger & Salisbury (3). Fat emboli were recognized as a major cause of collapse following injury by many authors including Dennis (4), who in his *Systems of Surgery* proposed the "Rule of 3's." Shock causes collapse in the first three hours after trauma, fat emboli cause collapse three days after injury, and pulmonary emboli three weeks after accident. Parenteral alcohol was the initial therapeutic agent used by Herrmann in 1932 (5) to treat fat embolism, as the lipolytic effects of this agent were thought to be beneficial.

INCIDENCE

Because of the subtle manifestations of the syndrome, its reported incidence has varied significantly. In the military experience, 10% of the battlefield deaths during World War I were thought to be secondary to pulmonary fat embolization, while the percentage increased markedly to 65% during World War II and was reported as 40% during the Korean conflict (6–8). Clinical reports indicate incidences of major fat embolization as high as 90% of the patients dying following automobile accidents (9). While there is no age or sex difference for individuals incurring a similar degree of multisystem injury, the percentage of male patients is higher because of more occupational and recreational exposure to major trauma.

PATHOPHYSIOLOGY OF FAT EMBOLISM

There is considerable controversy over both the source of the fat emboli and their mode of action. Two major theories exist. The first, known as the mechanical theory, was originally suggested by Gauss in 1924 (10). This theory proposes that the fat emboli arise by egress of local fat from the marrow at the site of fractures. These emboli are carried through the venous circulation to the pulmonary capillary bed, causing hemorrhagic or ischemic changes by temporary occlusion of the small vessels. Arnim & Grant (11) provided experimental support for the mechanical hypothesis by demonstrating bone-marrow emboli in pulmonary capillaries following long-bone fractures in animals. Others, while supporting the embolism and mechanical obstruction phase of the theory, feel that the emboli themselves may come from the coalescence of lipids mobilized from fat depots, particularly other than those located in marrow cavities (12).

The other, more popular theory has been called the physiochemical hypothesis and was first proposed by Lehmann & Moore in 1927 (3). This theory has evolved over a number of years to its present concept, which states that the pulmonary, cerebral, and cutaneous pathology is a result of the toxic effects of free fatty acids. The source of the free fatty acids may be either hydrolysis of neutral fat droplets by tissue lipase; the mobilization of fatty-acid stores by catecholamines, glucagon, and steroids; or both. Many researchers, including Peltier (14), have clearly demonstrated the toxic effects of free acids on lung tissue. These acidic substances cause disruption of the capillary alveolar membrane with alteration in surfactant activity leading to frothy edema, hemorrhage, and further loss of functional residual capacity. Intravenous injection of unbound free fatty acids, such as oleic or arachidonic acid, produce pulmonary decompensation in a variety of animal models. Kreis and associates (15) produced lethal hypoxia in dogs with intravenous injection of 0.07 ml/kg of oleic acid, with radiographic and pulmonary histologic changes consistent with the clinical picture of fat emboli syndrome. Slightly larger doses of oleic acid produced a marked decrease in pulmonary compliance and oxygen diffusion in a similar animal model (16). Incubation of the fatty acids with albumin prior to injection eliminated the toxic effects of these agents (17).

In the normal state, free fatty acids are transported through the intravascular space in a bound state, with albumin being the principal protein binder. Less than 1% of the circulating free fatty acids exists in the unbound state physiologically. Following trauma, there is a significant elevation in free fatty acid levels in all patients incurring serious injury (18). The degree of elevation appeared to be proportional to the magnitude of the injury. The source of the serum fatty acid appears to be increased lipolysis of adipose tissue. Carlson and his co-workers linked this excessive lipid mobilization with a microscopic pathological picture in animals consistent with clinical fat emboli in man (19). Recent clinical experience has demonstrated that patients who develop fat emboli syndrome have a significantly lower serum albumin than patients who remain asymptomatic, allowing significant portions of the free fatty acids to circulate unbound (20). Albumin administered intravenously during the early posttrauma period, to reduce the level of unbound free fatty acid, has produced a marked reduction in the incidence of fat emboli syndrome in patients with multiple injuries.

DIAGNOSIS

Clinical manifestations of fat emboli syndrome include dyspnea, confusion, and petechiae, since the target organs of this posttraumatic complication are the pulmonary, cerebral, and cutaneous systems. The initial respiratory components may be insidious, with slight tachypnea without marked distress. The syndrome can progress to severe dyspnea with cyanosis. Mental signs include anxiety, agitation, and confusion, progressing to coma in rare incidences. Careful examination of the torso may reveal a petechiae rash. Other areas where petechiae are often found include the conjunctivae, nail beds and uvula.

The peak incidence of this cause of pulmonary insufficiency is two to three days following injury. This temporal fact is useful in distinguishing fat embolism from other major causes of early respiratory distress. The manifestations of pulmonary contusion are usually evident in the first 24 hours following injury, while shock lung occurs later in the postinjury period.

The diagnosis of fat emboli syndrome is usually made on a clinical basis that includes a history of long-bone injury, shock, respiratory distress, progressive change in cerebral function, and the presence of petechial rash. However, there are some laboratory parameters that are confirmatory and aid the physician in the differential diagnosis. The chest X ray is usually normal at the time of admission. Findings of fluffy densities in both lung fields are not evident until about the second or third postinjury day. In patients with direct thoracic trauma-causing pneumohemothorax and pulmonary contusion, radiologic changes are unilateral and present within the first few hours following injury.

The value of fat globules in the urine as a criterion for fat emboli syndrome is questionable, since most patients following major trauma are noted to have urinary fat globules whether or not they develop respiratory distress (21). However, the absence of fat globules in patients with respiratory distress makes the diagnosis of fat emboli unlikely and points more toward shock lung. Shock lung causes respiratory distress that usually develops five days after injury, at which time fat is absent from the urine. Similar lack of diagnostic reliability has been reported for presence of fat in the sputum.

Electrocardiographic changes are not specific and indicate acute pulmonary hypertension and right ventricular strain. These electrocardiographic findings include the prominent S waves in Lead I, prominent Q waves in Lead III, varied S-T changes, and right-axis deviation.

Changes such as thrombocytopenia, hypofibrinogenemia, and prolongation of the partial thromboplastin times have been reported in patients with fat embolization and have been considered the cause of the clinical manifestations. However, controlled experimental work has demonstrated that these clotting abnormalities are secondary to hypovolemic shock rather than to fat embolization. Clinical experience in patients subjected to major trauma showed no statistical difference in platelet counts, fibrinogen levels, partial thromboplastin, and prothrombin times, in patients with and without respiratory distress (20).

Serum lipase and tributyrinase levels have been reported as valuable diagnostic aids for fat embolization. Both of these enzymes have been demonstrated to be elevated in patients with fat embolization, but they have not correlated well with the degree of severity of the clinical syndrome (22).

The arterial blood gas is the most helpful laboratory test in the diagnosis and management of fat embolization. An arterial PO_2 of less than 60 mm Hg on room air is indicative of significant oxygen desaturation and the need for respiratory support. As the syndrome initially develops, patients may be slightly alkalotic and hypocarbic secondary to hyperventilation; however, further deterioration results in continued hypoxia, hypercarbia, and acidosis.

TREATMENT

Many aspects of early treatment are thought to affect the development of fat emboli syndrome and proper attention to these details of care will minimize its incidence. Factors such as fracture instability, shock, and overresuscitation influence the incidence of this respiratory complication. All fractures should be stabilized as effectively and as early as possible using air splints at the scene of the accident, and balanced traction, casting, or internal stabilization in the hospital.

Prompt reversal of hypovolemic shock is extremely important in minimizing posttraumatic pulmonary insufficiency. Use of electrolyte solutions and whole blood for rapid volume expansion is an important factor. While estimation of blood loss is often difficult in this patient population because of continued, hidden blood losses, careful attention to serial circumferential measurements of injured limbs, the use of central venous-pressure monitoring in unstable patients, and accurately maintained intake and output balances will minimize fluid overload and pulmonary edema.

While the value of colloid during the initial resuscitation as a volume expander has been questioned, the use of albumin replacement for its fatty-acid binding capabilities, by maintaining a serum albumin level greater than 3.5%, has been valuable in reducing the incidence of pulmonary insufficiency from fat emboli in our clinical experience.

Historically, many agents have been used to treat fat emboli syndrome including 5% ethyl alcohol by Herrmann in 1932 (5). He used this agent after it had been shown to inhibit lipase activity both in vitro and in vivo. However, continued use of this agent has never been demonstrated to be effective in reducing pulmonary insufficiency of fat embolism once it has developed. Heparin is another agent that has been advocated because of its lipemic clearing capability and anticlotting properties. Nonetheless, because of the increased bleeding problems associated with trauma, heparin has not been extensively used for this syndrome.

The most effective agent in treating fat emboli syndrome once it has developed has been systemic corticosteroids. Ashbaugh & Petty (23) first reported its use and effectiveness in reversing respiratory failure with massive fat emboli. High-dose steroids have been demonstrated to reduce the high mortality rate associated with this disease. The major action of corticosteroids in reversing the toxic effect of free fatty acids appears to be their anti-inflammatory property. Experimental studies with fat embolization have shown that corticosteroids reduce hypoxia secondary to this process.

Initially the hypoxia can be reversed with the use of nasal oxygen; however, since rapid fall in the Po_2 may develop in the patients over a period of 12 to 18 hours, serial blood-gas determinations are necessary. If arterial Po_2 greater than 70 mm Hg cannot be maintained on nasal oxygen, intubation and ventilatory support with a volume ventilator may be necessary. High tidal volumes in the range of 15 ml/kg body weight have been effective in stabilizing and reversing pulmonary posttraumatic respiratory distress. The addition of positive end-expiratory pressure has been shown to be beneficial in increasing the functional residual capacity and decreasing pulmonary shunting in patients (24).

SUMMARY

Fat emboli syndrome is one of the major causes of early posttraumatic respiratory distress. Careful attention to the multiple factors associated with the early resuscitative period can minimize this lethal complication of multisystem injury. The use of high-dose steroids and appropriate ventilatory support will decrease the high mortality rate associated with this disease syndrome once it has developed.

Literature Cited

1. Zenker, F. A. 1861. *Beitrage zur Anatomie und Physiologie der Lunge.* Dresden: Braunsdorf. 20 pp.
2. Bergmann, E. B. 1873. Ein Fall todlicher fetlenbolic. *Berl. Klin. Wochenschr.* 10:385–87
3. Fenger, D., Salisbury, J. H. 1879. Diffuse multiple capillary fat embolism in the lungs and brain as a fatal complication in common fracture: illustrated by a case. *Chicago Med. J. Exam.* 39:587–95
4. Dennis, F. S. 1895. *System of Surgery,* Vol. 1. Philadelphia: Lea Bros. 533 pp.
5. Herrmann, L. G. 1932–1933. Effect of dextrose alcohol mixture upon pulmonary fat embolism. *Proc. Soc. Exp. Biol. Med.* 30:558–59
6. Malloy, T. B., Sullivan, E. R., Burnett, C. H. 1950. VII. The general pathology of traumatic shock. *Surgery* 27:629–44
7. Scully, R. E. 1956. Fat embolism in Korean battle casualties; its incidence, clinical significance and pathologic aspects. *Am. J. Pathol.* 32:379–403
8. Sutton, G. E. 1918. Pulmonary fat embolism and its relation to traumatic shock. *Br. Med. J.* 2:368–70
9. Greendyke, R. M. 1964. Fat embolism in car deaths, New York. *J. Forensic Sci.* 9:201–8
10. Gauss, H. 1924. The pathology of fat embolism. *Arch. Surg.* 9:593–605
11. Arnim, J., Grant, R. E. 1966. Observations on gross pulmonary fat embolism in man and in the rabbit. *Can. J. Surg.* 9:286–93
12. Bergenty, S. E. 1961. Studies on the genesis of post traumatic fat embolism. *Acta Clin. Scand.* 282:1–72
13. Lehman, E. P., Moore, R. M. 1927. Fat embolism including experimental pro-

duction without trauma. *Arch. Surg.* 14:621–62

14. Peltier, L. F. 1956. Fat embolism: The toxic properties of neutral fat and free fatty acids. *Surgery* 40:665–70
15. Kreis, W. R., Lindenaur, S. M., Dent, T. L. 1973. Corticosteroids in experimental fat embolization. *J. Surg. Res.* 14:238–46
16. Cahill, J. M., Daly, B. F. T., Byrne, J. J. 1974. Ventilatory and circulatory response to oleic acid embolus. *J. Trauma* 14:73–76
17. Mays, E. T., Schrodt, R., Chung, R. 1976. Role of protein binding of cytotoxic anions in post-traumatic pulmonary insufficiency. *J. Trauma* 16: 290–99
18. McNamara, J. D., Molat, M., Dunn, R., Burran, E. L., Stremple, J. F. 1972. Lipid metabolism after trauma. *J. Thorac. Cardiovasc. Surg.* 66:968–72
19. Carlson, L. A. 1970. Mobilization and utilization of lipids after trauma: Relation to caloric homeostasis. Ciba Found. Symp: *Energy Metabolism in Trauma,* pp. 155–73. London: Churchill
20. Moylan, J. A., Evenson, M. E., Birnbaum, M., Katz, A. 1976. Fat emboli syndrome. *J. Trauma* 16:341–47
21. Mortón, K. S. 1956. Fat embolism: Incidence of urinary fat in trauma. *Can. Med. Assoc. J.* 74:441–48
22. Herndon, J. H., Riseborough, E. J., Fischer, J. E. 1971. Fat embolism, a review of current concepts. *J. Trauma* 11:673–80
23. Ashbaugh, D. G., Petty, T. L. 1966. The use of corticosteroids in the treatment of respiratory failure associated with massive fat embolism. *Surg. Gynecol. Obstet.* 123:493–500
24. Ashbaugh, D. G., Petty, T. L. 1973. Positive end-expiratory pressure, physiology, indications and contraindications. *J. Thorac. Cardiovasc. Surg.* 65:165–70

Ann. Rev. Med. 1977. 28:91–102
Copyright © 1977 by Annual Reviews Inc. All rights reserved

GLOMERULAR FILTRATION OF FLUID AND MACROMOLECULES: THE RENAL RESPONSE TO INJURY

❖7226

David A. Maddox, Ph.D.
Departments of Medicine and Physiology and the Cardiovascular Research Institute
of the University of California, San Francisco, California 94143

Barry M. Brenner, M.D.
Laboratory of Kidney and Electrolyte Physiology, Department of Medicine,
Peter Bent Brigham Hospital, Boston, Massachusetts 02115

INTRODUCTION

Despite significant advances in the understanding of the immunological and morphological aspects of a variety of disorders involving the kidney in which the glomerular filtration of water and macromolecules is affected, there has been, until recently, relatively little information concerning the mechanisms responsible for these changes in filtration. This is due in large part to the fact that glomeruli are seldom encountered as surface structures in the mammalian kidney, and are therefore inaccessible to direct study in vivo. This restriction has been overcome, however, with the recent discovery of a unique strain of Wistar rats (the so-called Munich-Wistar strain) with surface glomeruli. This advantage, together with technological and methodological developments allowing the direct assessment of hydraulic and oncotic pressures in the glomerular and periglomerular microcirculation (1, 2), has made possible an analysis of the determinants of fluid and solute filtration in this mammalian species. This review summarizes our current understanding of the transglomerular transport of fluid and macromolecules, both in the normal state and in response to renal injury.

DETERMINANTS OF GLOMERULAR ULTRAFILTRATION

Normal Glomerular Dynamics

The rate of production of a nearly ideal ultrafiltrate of plasma across the glomerular capillary walls, as in other capillary beds, is governed by the imbalance between the transcapillary hydraulic and colloid osmotic pressures. At any point along a glomerular capillary the local rate of ultrafiltration, J_v, equals $k \cdot (\Delta P - \Delta \pi)$, where k is the effective hydraulic permeability of the capillary wall, and ΔP and $\Delta \pi$ are the transcapillary hydraulic and colloid osmotic-pressure differences, respectively. ΔP is the difference between glomerular capillary hydraulic pressure, P_{GC}, and Bowman's space hydraulic pressure, P_T, while $\Delta \pi$ refers to the corresponding colloid osmotic-pressure difference between the glomerular capillaries (π_{GC}) and Bowman's space (π_T). Under normal hydropenic conditions P_{GC} has been found to average approximately 45 mm Hg in both the Munich-Wistar rat (3–9) and the squirrel monkey (10), while P_T averages approximately 10 mm Hg. $\overline{\Delta P}$, the difference between these values of P_{GC} and P_T averaged along the length of the glomerular capillary, is therefore equal to approximately 35 mm Hg. In the rat, $\Delta \pi$, equal to π_{GC} since π_T is negligible (11), increases from a value of about 20 mm Hg at the afferent end (π_A) of the capillary network to approximately 35 mm Hg at the efferent end (π_E). The net local driving force for ultrafiltration, P_{UF}, therefore declines from a maximum value of about 15 mm Hg at the afferent end of the glomerulus to zero at the efferent end. Thus, by the efferent end of the capillary network, $\Delta \pi$ rises to a value that, on the average, exactly equals ΔP (3–10). This equality of ΔP and $\Delta \pi$ is referred to as *filtration pressure equilibrium*.

Typical profiles of ΔP and $\Delta \pi$ for the glomerular capillary of normal hydropenic animals are illustrated in Figure 1, where the abscissa represents the fractional distance along an idealized glomerular capillary, 0 being the afferent end and 1 the efferent end. Precise determinations of π can only be made on samples of systemic plasma (assumed to be equal to afferent arteriolar plasma) and plasma from efferent arterioles, thereby precluding direct determination of the exact profile of the change in $\Delta \pi$ with distance along the capillary. Thus $\Delta \pi$ curves A and B in the left and center panels of Figure 1 both satisfy the experimental finding in normal hydropenia of filtration-pressure equilibrium being achieved by the efferent end of the glomerular capillaries. Importantly, however, curves A and B are but two of an infinite number of profiles consistent with the measurements of π_A and π_E obtained in

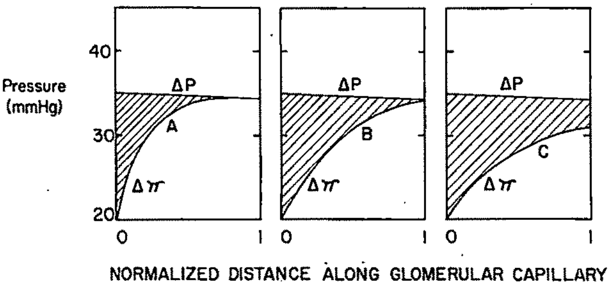


Figure 1 Comparison of transcapillary hydraulic (ΔP)- and oncotic ($\Delta \pi$)-pressure profiles along idealized glomerular capillaries under hypothetical conditions of filtration pressure equilibrium (*left and center panels*) or disequilibrium (*right panel*).

hydropenia. Hence the precise point at which $P_{UF} = 0$ cannot be determined under equilibrium conditions.

For the glomerulus as a whole, the single-nephron glomerular filtration rate (SNGFR) is the product of the ultrafiltration coefficient K_f, and the net driving pressure averaged over the length of the capillary, $\bar{P}_{UF}$. $\bar{P}_{UF}$ is the difference between the mean transcapillary hydraulic- and oncotic-pressure differences, $\overline{\Delta P}$ and $\overline{\Delta \pi}$, respectively. The shaded areas in Figure 1 (that is, the areas between the ΔP and $\Delta \pi$ curves) correspond to the value of $\bar{P}_{UF}$ in each panel. K_f is the product of the effective hydraulic permeability of the capillary wall and the surface area available for filtration. When filtration-pressure equilibrium exists it is not possible to estimate $\bar{P}_{UF}$, and hence K_f, due to the uncertainty in determining the exact $\Delta \pi$ profile.

A variety of studies, designed to examine how variations in glomerular transcapillary pressures result in changes in SNGFR, have revealed that such changes in SNGFR result primarily from variations in the rate of rise of $\Delta \pi$, rather than from large changes in $\overline{\Delta P}$ (4, 5, 7, 9, 12). Alterations in the rate of rise of $\Delta \pi$ result from variations in the rate of glomerular plasma flow (Q_A), with increases in Q_A tending to displace the point at which filtration-pressure equilibrium is reached further toward the efferent end of the glomerular capillary (13), as from curve A in the left-hand panel of Figure 1 to curve B in the center panel. As a result, $\bar{P}_{UF}$ increases, as denoted by the increase in the shaded area, leading to an increase in SNGFR. Decreases in Q_A have the opposite effect. It has been concluded from these observations that SNGFR is highly plasma flow–dependent, at least in the Munich-Wistar rat.

Since it is impossible, under conditions of filtration-pressure equilibrium, to distinguish among the many $\Delta \pi$ profiles consistent with equilibrium, only a minimum value of K_f can be estimated. If, however, Q_A is increased to levels above normal so that $\Delta \pi$ no longer rises to a value large enough to equal ΔP, as in the right-hand panel of Figure 1, a condition of filtration-pressure disequilibrium will obtain. Under these conditions, a unique $\Delta \pi$ profile can be computed from the measured values of afferent and efferent protein concentration, ΔP, and Q_A, using the mathematical model of glomerular filtration derived by Deen, Robertson & Brenner (13).

This approach of increasing Q_A to levels high enough to ensure disequilibrium in order to estimate the value of K_f in the Munich-Wistar rat has been achieved both by intravenous infusion of isoncotic rat plasma (12) and by isovolemic reduction in hematocrit (7). In both instances K_f was found to average about 4.8 nl·(min·mm Hg)$^{-1}$. Furthermore, K_f remained essentially unchanged over a twofold range of changes in Q_A (12), suggesting that K_f does not vary with Q_A.

Response to Reduced Renal Mass

Reduction in functional renal mass, when the result either of partial surgical ablation or a primary nonglomerular disease such as pyelonephritis, generally leads to an adaptive increase in the rate of glomerular filtration in surviving nephrons (5, 14–19). To ascertain the basis for this increase in GFR, Deen et al (5) recently examined glomerular dynamics in uninephrectomized Munich-Wistar rats studied 2–4 weeks following surgery. Characteristic of compensatory hypertrophy, both total GFR and the total renal mass of the contralateral kidney increased by approximately 50%. Results of renal micropuncture studies in these animals and in a separate group of normal hydropenic (sham-operated) control rats are shown in Figure 2. Along with total GFR in the remaining kidney, both SNGFR and Q_A increased proportionately following uninephrectomy, with single-nephron filtration fraction (SNFF) remaining relatively constant. Despite no signficant differences in systemic arterial pressure ($\overline{AP}$) between the two groups, $\overline{\Delta P}$ was significantly higher following uninephrectomy, averaging about 40 mm Hg, compared to 34 mm Hg in controls (Figure 2). This was entirely the consequence of an increase in $\overline{P}_{GC}$, P_T remaining unchanged. π_A was not different between the two groups and filtration-pressure equilibrium was observed in both groups of rats.

It can be seen from Figure 2 that the increase in SNGFR following uninephrectomy could have resulted from an increase in K_f, $\overline{P}_{UF}$, or both. As discussed previously, a unique value of K_f cannot be determined under conditions of filtration pressure equilibrium. To estimate K_f for uninephrectomized rats, Deen et al (5) produce disequilibrium by isoncotic plasma expansion, which had the desirable effect of increasing Q_A. Under these conditions K_f was found to average 4.7 nl $(\text{min}\cdot\text{mmHg})^{-1}$, a value similar to that found in non-nephrectomized control rats (7, 12). Since π_A and K_f were essentially unchanged following uninephrectomy,

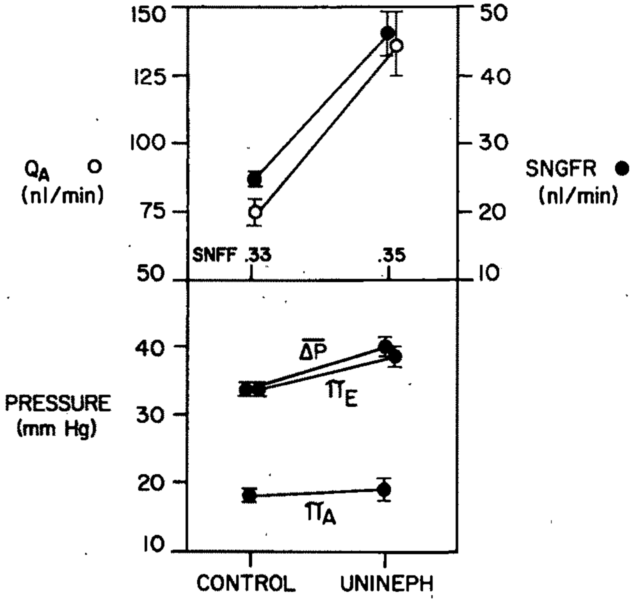


Figure 2 Summary of effects of uninephrectomy on some measured determinants of glomerular ultrafiltration. Data (Mean ± SE) were obtained from 8 control Munich-Wistar rats (*left*) and 10 rats (*right*) subjected to uninephrectomy 2–4 weeks prior to study. Reproduced from (41) with permission.

the observed increases in the mean net ultrafiltration pressure, $\overline{P}_{UF}$ and resulting increase in SNGFR must have been the consequence of the increases in Q_A and $\overline{\Delta P}$. An estimation of the individual contribution of these changes in Q_A and $\overline{\Delta P}$ to the observed changes in SNGFR was calculated (13) and revealed that 70% of the compensatory increase in SNGFR was due to the rise in Q_A, the remainder being due to the increase in $\overline{\Delta P}$.

In accord with these findings, Kaufman et al (18) observed that a 50% reduction in renal mass in the rat resulted in proportional increases in SNGFR and Q_A, with SNFF remaining the same as in non-nephrectomized control rats. Following 75% ablation of renal mass, however, SNGFR and Q_A increased further, but with a decrease in SNFF. Assuming that there was no substantial fall in $\overline{\Delta P}$ or increase in π_A in the animals studied by Kaufman et al (18), the further increase in Q_A to remaining nephrons probably produced filtration-pressure disequilibrium. Thus it appears that following 50% or even 75% reduction in renal mass, the compensatory increases in SNGFR are primarily governed by the extent to which Q_A increases in surviving nephrons. Here again, the results accord well with the earlier conclusion that GFR is highly plasma flow-dependent.

Response to Primary Glomerular Injury

In contrast to the observed increases in SNGFR, which tend to minimize the degree of reduction in total GFR that would otherwise occur when renal mass is reduced by surgical ablation, progressive destruction of glomerular capillaries in a variety of glomerulopathic states is generally attended by reductions in both whole-kidney and single-nephron GFR. Relatively little is known of the forces governing glomerular ultrafiltration in these glomerulopathies, despite considerable advances in our understanding of the morphological and immunological aspects of these disorders.

Maddox et al (6), however, have recently examined the forces determining glomerular filtration in Munich-Wistar rats 1–2 weeks following induction of a mild to moderate form of nephrotoxic serum nephritis (NSN). Both light and electron microscopy revealed that the proliferative inflammatory response associated with this form of glomerular injury caused partial or complete obliteration of some capillary loops, while other capillaries appeared essentially normal. Despite this morphologic evidence of widespread glomerular injury, glomerular filtration rates (single-nephron and whole-kidney) and glomerular plasma-flow rates (Q_A) were not different, on the average, from values in non-nephritic control rats, as shown in Figure 3 (*top panel*). In NSN, $\overline{\Delta P}$ was elevated considerably above normal, due solely to increases in $\overline{P}_{GC}$. Since Q_A and π_A remained essentially constant, the failure of SNGFR to rise in response to the increase in $\overline{\Delta P}$ in NSN rats was the consequence of a concomitant reduction in K_f, to a value averaging one third of normal. Allison, Wilson & Gottschalk (20) have similarly observed reduction in K_f in studies of more severe NSN as well as Heymann's nephritis in the rat. It is likely that the fall in K_f observed in NSN rats is at least partly the result of the reduction in the number of patent capillaries in each glomerulus, i.e. a reduction in surface area available for ultrafiltration. Whether a decrease in effective hydraulic permeability also contributes to the fall in K_f cannot be ascertained from morphological considerations. This

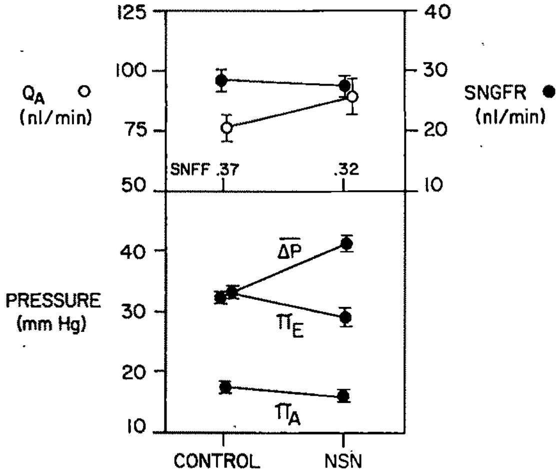


Figure 3 Summary of some of the determinants of SNGFR in 16 control Munich-Wistar rats (*left*) and 22 rats with nephrotoxic serum nephritis (NSN, *right*). Values are expressed as means ± SE. Reproduced from (41) with permission.

latter possibility is suggested, however, by the recent findings of Blantz & Wilson (21), who observed a fall in K_f one hour after intravenous injection of nephrotoxic serum. At this early interval following injection of nephrotoxic serum, the number of patent glomerular capillaries was not found to be reduced appreciably, so that a fall in hydraulic permeability, rather than a reduction in surface area, was postulated to account for their observed marked fall in K_f.

In the study of Maddox et al (6), in the early stage of NSN, SNGFR and total-kidney GFR remained at normal levels despite a fall in K_f to one third of normal, because of the compensating effect of an increase in $\overline{\Delta P}$. If $\overline{\Delta P}$ remains elevated in chronic glomerulonephritis, as suggested by the findings of Allison et al (20), the progressive deterioration in filtration rate characteristically noted in more advanced stages of this disorder must result from further reductions in K_f and/or declines in Q_A.

PERMSELECTIVITY OF THE GLOMERULAR CAPILLARY WALL

Studies in Normal Animals

The glomerular capillary wall normally possesses the ability to discriminate among molecules of varying size. While molecules the size of inulin or smaller normally permeate the capillary wall without measurable restriction, the transglomerular passage of molecules of increasingly greater size diminishes progressively, the extent of restriction being almost complete for molecules the size of serum albumin or greater.

Our current views of the mechanisms governing this permselectivity of the glomerular capillary wall derive largely from differential solute-clearance studies, in which the urinary excretion of some test macromolecule (such as dextran) is compared to that of a reference solute (such as inulin) that appears in Bowman's space in the same concentration as in plasma water. If both test and reference solutes are neither secreted nor reabsorbed, fractional clearance determined in this manner is equivalent to the ratio of the concentration of the test macromolecule in Bowman's space to that in plasma water.

The dashed line in Figure 4 shows results of measurements by Chang and coworkers (9) of the transport of uncharged, tritiated dextrans (D) of various sizes across glomerular capillaries in the normal hydropenic Munich-Wistar rat. Fractional clearances of D are plotted as a function of effective radius of D, a value of one on the ordinate corresponding to a clearance of D equal to that of inulin. As shown, measurable restriction to the transport of D (i.e. fractional clearance ratios less than one) does not occur until the effective radius exceeds approximately 21 Å. For larger molecular radii, fractional clearances of D decrease progressively with increasing size, approaching zero at radii greater than 42 Å. Similar fractional solute-clearance profiles have been reported in other experimental animals and man (22–25).

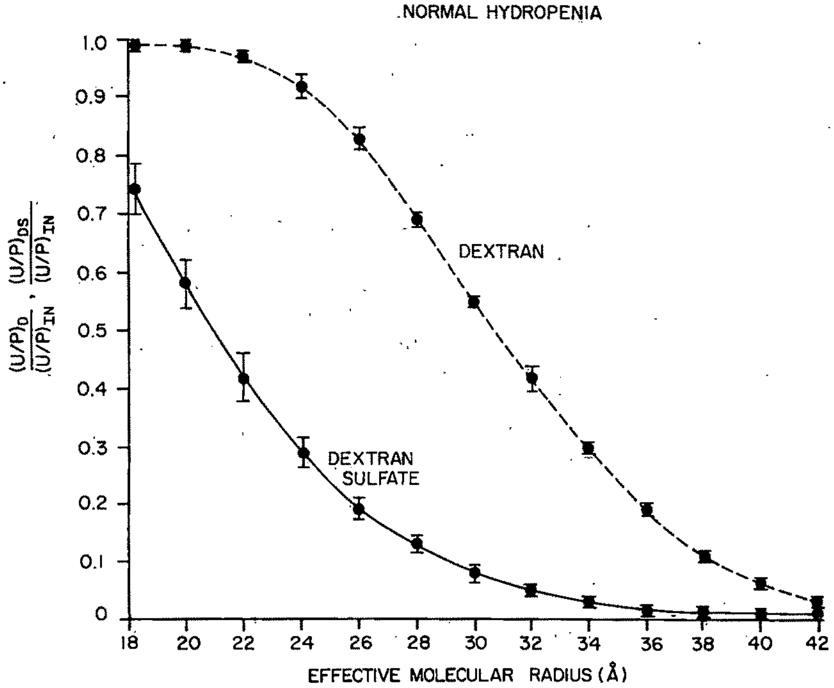


Figure 4 Fractional clearance of dextran sulfate (DS) and neutral dextran (D), plotted as a function of effective molecular radius. Values are expressed as means ± SE. Reproduced from (29) with permission.

It is important to appreciate that solute and water transport across the glomerular capillary wall are closely coupled, with the Bowman's space-to-plasma concentration of a macromolecule being inversely dependent on GFR (24–27). Accordingly, the transglomerular transport of macromolecules will be influenced not only by the physicochemical properties of the capillary wall, but also by the other determinants of glomerular ultrafiltration, namely the transcapillary hydraulic- and oncotic-pressure differences and the glomerular capillary plasma-flow rate, Q_A. Recently Chang et al (9) compared fractional clearances of D in Munich-Wistar rats before and during volume expansion with isoncotic plasma, the main effect of the latter being to increase Q_A. In keeping with theoretical predictions that the transport of a macromolecule increases less than in proportion to the increase in the glomerular filtration of water (25–28), the measured threefold increases in Q_A during plasma loading resulted in falls in fractional clearances of D. From these results, Chang et al (9) determined that dextran filtration is influenced by diffusion as well as convection.

In addition to the effects of molecular size and the various determinants of GFR, fractional clearances of macromolecules have recently been shown to be influenced by electrical factors as well. To illustrate this latter influence, recall that the Bowman's space-to-plasma concentration ratio for albumin is normally very low, averaging less than 0.001 in the normal rat (11). The fractional clearance of a neutral dextran molecule of effective radius equivalent to albumin, ∼36 Å, averages 0.19 (Figure 4), or some two orders of magnitude greater than that of albumin. That this difference between the clearance of albumin and that of nonprotein macromolecules is not due primarily to albumin reabsorption by the renal tubule is indicated by the observation that the difference persists even when fractional clearances are measured using fluid obtained from Bowman's space (9). Since albumin is a polyanion, Chang et al (29) examined the effect of molecular electrical charge on glomerular permselectivity by studying the fractional clearances of dextran sulfate, (DS), anionic polymers structurally similar to neutral D. As shown in Figure 4, whereas a measured restriction to the transglomerular passage of neutral D does not occur until the effective radius exceeds 21 Å, restriction to DS is evident over the entire range of molecular radii studied. It can be seen that for DS with an effective radius of 36 Å, fractional clearance averaged 0.015, thus approaching that of albumin. These differences in the transport of DS vs D suggest electrostatic repulsion of charged macromolecules by some component(s) of the glomerular capillary wall (see below).

Studies in Animals With Primary Glomerular Injury

The fractional-clearance profiles obtained with macromolecules in both normal and pathological states are often erroneously taken to reflect the permselective properties of the glomerular capillary wall alone. As indicated above, however, these fractional-clearance profiles are likely to be influenced not only by the properties of the capillary wall, but also by the determinants of glomerular filtration rate and by the net charge of the molecule under study. Thus, when examining macromolecular transport across glomerular capillaries in studies of primary glomerular injury, it is essential to measure the pertinent driving forces and flows governing ultrafiltration and to define the physicochemical properties of the test substance. Chang et al (30) recently examined the transport of neutral D across accessible glomerular capillaries in Munich-Wistar rats following induction of nephrotoxic serum nephritis. Figure 5 (*open circles*) shows the relationship between fractional clearances of D and effective radius of D for 8 NSN rats studied during hydropenia. Whereas in non-nephritic control rats (*solid circles*), restriction to transport of D does not occur until the effective radius exceeds 21 Å, in NSN rats restriction is evident over the entire size range of D examined. Since there is no measurable restriction to the transport of inulin (~14 Å) in NSN rats (30), these findings indicate that restriction to D in NSN becomes manifest at some radius above 14 Å but below 18 Å.

Although SNGFR was normal, on the average, in these NSN rats, SNFF was lower than that observed in non-nephritic control rats, despite values for $\overline{\Delta P}$ in excess of normal (30). This resulted from a marked fall in K_f induced by glomerular injury, as discussed earlier. Curve A of Figure 5 depicts the effects on fractional clearances of D expected solely from the measured changes in these glomerular pressures and flows in NSN rats (30). Since these changes in the determinants of SNGFR alone are insufficient to account for the observed marked displacement of

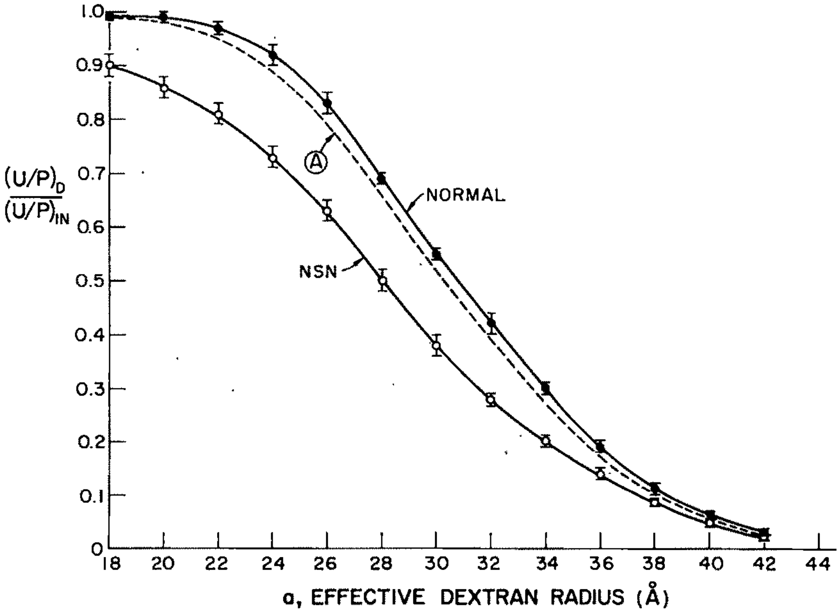


Figure 5 Fractional dextran clearances plotted as a function of effective dextran radius for NSN rats (*open circles*) and for a group of non-nephritic normal hydropenic control rats (*solid circles*). Values are expressed as means ± SE. The significance of curve *A* is discussed in the text. Reproduced from (30) with permission.

the clearance profile of D in NSN, changes in the permselectivity properties of the glomerular capillary wall must have occurred in NSN. Of interest was the observation that the excretion of albumin was found to increase in these NSN rats, whereas the excretion of D having the same effective radius as albumin (~36 Å) decreased.

To examine the possibility of disease-induced alterations of the fixed electrical charges residing on the walls of the glomerular capillaries in NSN rats, the effect of which might be to increase the filtration of polyanionic, but not neutral, macromolecules, Bennett et al (31) examined the transport of DS in NSN rats. As shown in Figure 6, fractional clearances of DS were substantially greater, at any given molecular size, than those found in normal Munich-Wistar rats. These findings suggest that loss of fixed negative charges from the glomerular capillary wall may be a major cause of increased filtration of plasma proteins following experimental or spontaneous glomerular injury. In this regard, it has been known for some time that the glomerular epithelial cells and their foot processes are covered with a thin layer of an anionic glycoprotein (sialoprotein) (32–35). In addition, the glomerular basement membrane (GBM) has been shown to contain sialic acid (36, 37). These sialoprotein layers may normally serve to restrict the passage of anionic mac-

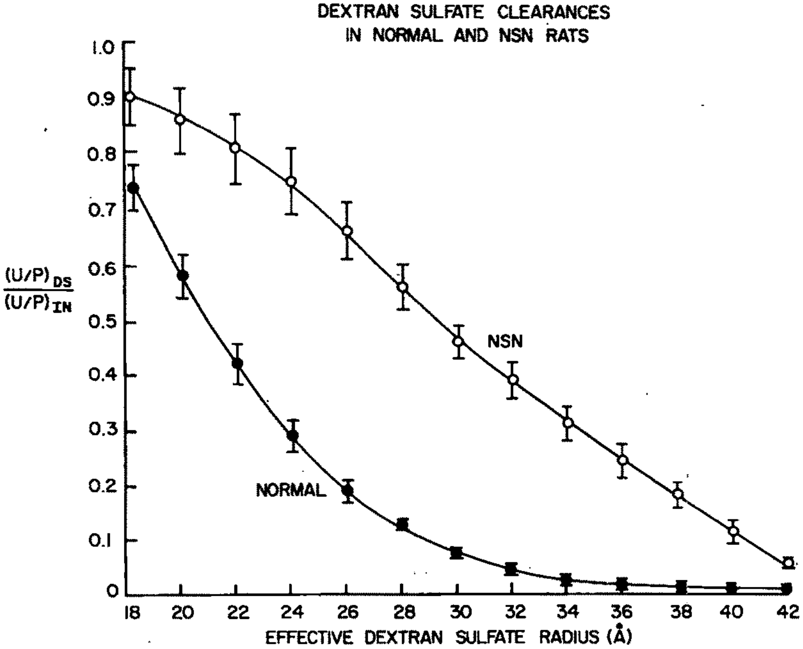


Figure 6 Fractional dextran sulfate (DS) clearances plotted as a function of effective DS radius for NSN rats (*open circles*) and for a group of non-nephritic normal hydropenic control rats (*solid circles*). Values are expressed as means ± SE. Reproduced from (31) with permission.

romolecules such as plasma proteins (32) and DS. In support of this possibility, glomerular sialoprotein content has been found to be diminished in glomerulopathies associated with proteinuria (31, 34, 38). Moreover, the sialic-acid content of isolated GBM is reduced in a variety of experimental and clinical proteinuric states (37, 39, 40). Obviously, it remains to be determined whether a causal relationship between loss of glomerular polyanion and proteinuria exists in man. If so, therapeutic efforts should then be directed at restoring fixed negative charges to the glomerular capillary wall.

Literature Cited

1. Wiederhielm, C. A., Woodbury, J. W., Kirk, S., Rushmer, R. F. 1964. Pulsatile pressures in the microcirculation of frog's mesentery. *Am. J. Physiol.* 207:173–76
2. Brenner, B. M., Falchuk, K. H., Keimowitz, R. I., Berliner, R. W. 1969. The relationship between peritubular capillary protein concentration and fluid reabsorption by the renal proximal tubule. *J. Clin. Invest.* 48:1519–31
3. Brenner, B. M., Troy, J. L., Daugharty, T. M. 1971. The dynamics of glomerular ultrafiltration in the rat. *J. Clin. Invest.* 50:1776–80
4. Brenner, B. M., Troy, J. L., Daugharty, T. M., Deen, W. M., Robertson, C. R. 1972. Dynamics of glomerular ultrafiltration in the rat. II. Plasma-flow dependence of GFR. *Am. J. Physiol.* 223:1184–90
5. Deen, W. M., Maddox, D. A., Robertson, C. R., Brenner, B. M. 1974. Dynamics of glomerular ultrafiltration in the rat. VII. Reponse to reduced renal mass. *Am. J. Physiol.* 277:556–62
6. Maddox, D. A. Bennett, C. M., Deen, W. M., Glassock, R. J. Knudsen, D., Daugharty, T. M., Brenner, B. M. 1975. Determinants of glomerular filtration in experimental glomerulonephritis in the rat. *J. Clin. Invest.* 55:305–18
7. Myers, B. D., Deen, W. M., Robertson, C. R., Brenner, B. M. 1975. Dynamics of glomerular ultrafiltration in the rat. VIII. Effects of hematocrit. *Circ. Res.* 36:425–35
8. Blantz, R. C. 1974. Effect of mannitol in glomerular ultrafiltration in the hydropenic rat. *J. Clin. Invest.* 54:1135–43
9. Chang, R. L. S., Ueki, I. F., Troy, J. L., Deen, W. M., Robertson, C. R., Brenner, B. M. 1975. Permselectivity of the glomerular capillary wall to macromolecules. II. Experimental studies in rats using neutral dextran. *Biophys. J.* 15:887–906
10. Maddox, D. A., Deen, W. M., Brenner, B. M. 1974. Dynamics of glomerular ultrafiltration. VI. Studies in the primate. *Kidney Int.* 5:271–78
11. Eisenbach, G. M., Van Liew, J. B., Boylan, J. W. 1975. Effect of angiotensin on the filtration of protein in the rat kidney: A micropuncture study. *Kidney Int.* 8:80–87
12. Deen, W. M., Troy, J. L., Robertson, C. R., Brenner, B. M. 1973. Dynamics of glomerular ultrafiltration in the rat. IV. Determination of the ultrafiltration coefficient. *J. Clin. Invest.* 52:1500–8
13. Deen, W. M., Robertson, C. R., Brenner, B. M. 1972. A model of glomerular ultrafiltration in the rat. *Am. J. Physiol.* 223:1178–83
14. Allison, M. E. M., Lipham, E. M., Lassiter, W. E., Gottschalk, C. W. 1973. The acutely reduced kidney. *Kidney Int.* 3:354–63
15. Bank, N., Aynedjian, H. S. 1966. Individual nephron function in experimental pyelonephritis. I. Glomerular filtration rate and proximal tubular sodium, potassium, and water reabsorption. *J. Lab. Clin. Med.* 68:713–37
16. Gottschalk, C. W. 1971. Function of the chronically diseased kidney. The adaptive nephron. *Circ. Res.* 23, 24, Suppl. II: 1–13
17. Hayslett, J. P., Kashgarian, M., Epstein, F. H. 1968. Functional correlates of compensatory renal hypertrophy. *J. Clin. Invest.* 47:774–82
18. Kaufman, J. M., Di Meola, H. J., Siegel, N. J., Lytton, B., Kashgarian, M., Hayslett, J. P 1974. Compensatory adaptation of structure and function following progressive renal ablation. *Kidney Int.* 6:10–17
19. Weber, H., Lin, K-Y., Bricker, N. S. 1975. Effect of sodium intake in single nephron glomerular filtration rate and sodium reabsorption in experimental uremia. *Kidney Int.* 8:14–20

20. Allison, M. E. M., Wilson, C. B., Gottschalk, C. W. 1974. Pathophysiology of experimental glomerulonephritis in rats. *J. Clin. Invest.* 53:1402–23
21. Blantz, R. C., Wilson, C. B. 1975. Glomerular filtration following acute immunological injury. *Am. Soc. Nephrol.* 8:60 (Abstr.)
22. Arturson, G., Groth, T., Grotte, G. 1971. Human glomerular membrane porosity and filtration pressure: Dextran clearance data analyzed by theoretical models. *Clin. Sci.* 40:137–58
23. Gassee, J. P. 1973. Effect of acetylcholine on glomerular sieving of macromolecules. *Pfluegers Arch.* 342: 239–54
24. Lambert, P. P., Dubois, R., Decoodt, P., Gassee, J. P., Verniory, A. 1975. Determination of glomerular intracapillary and transcapillary pressure gradients from sieving data. II. A physiological study in the normal dog. *Pfluegers Arch.* 359:1–22
25. Verniory, A., Dubois, R., De Coodt, P., Gassee, J. P., Lambert, P. P. 1973. Measurement of the permeability of biological membranes. Application to the glomerular wall. *J. Gen Physiol.* 62:489–507
26. Pappenheimer, J. R. 1953. Passage of molecules through capillary walls. *Physiol. Rev.* 33:387–423
27. Renkin, E. M. 1954. Filtration, diffusion and molecular sieving through porous cellulose membranes. *J. Gen. Physiol.* 38:225–43
28. Chang, R. L. S., Robertson, C. R., Deen, W. M., Brenner, B. M. 1975. Permselectivity of the glomerular capillary wall to macromolecules. I. Theoretical considerations. *Biophys. J.* 15:861–86
29. Chang, R. L. S., Deen, W. M., Robertson, C. R., Brenner, B. M. 1975. Permselectivity of the glomerular capillary wall. III. Restricted transport of polyanions. *Kidney Int.* 8:212–18
30. Chang, R. L. S., Deen, W. M., Robertson, C. R., Bennett, C. M., Glassock, R. J., Brenner, B. M. 1976. Permselectivity of the glomerular capillary wall. Studies of experimental glomerulonephritis in the rat using neutral dextran. *J. Clin. Invest.* 57:1272–86
31. Bennett, C. M., Glassock, R. J., Chang, R. L. S., Deen, W. M., Robertson, C. R., Brenner, B. M. 1976. Permselectivity of the glomerular capillary wall. Studies of experimental glomerulonephritis in the rat using dextran sulfate. *J. Clin. Invest.* 57:1287–94
32. Mohos, S. C., Skoza, L. 1969. Glomerular sialoprotein. *Science* 164:1519–21
33. Mohos, S. C., Skoza, L. 1970. Histochemical demonstration and localization of sialoproteins in the glomerulus. *Exp. Mol. Pathol.* 12:316–23
34. Michael, A. F., Blau, E., Vernier, R. L., 1970. Glomerular polyanion: Alteration in aminonucleoside nephrosis. *Lab. Invest.* 23:649–57
35. Latta, H., Johnson, W. H., Stanley, T. M. 1975. Sialoglycoproteins and filtration barriers in the glomerular capillary wall. *J. Ultrastruct. Res.* 51:354–76
36. Kefalides, N. A. 1973. Structure and biosynthesis of basement membranes. *Int. Rev. Connect. Tissue Res.* 6:63–104
37. Lui, S., Kalant, N. 1974. Carbohydrate of the glomerular basement membrane in normal and nephrotic rats. *Exp. Mol. Pathol.* 21:52–62
38. Blau, E. B., Haas, J. E. 1973. Glomerular sialic acid and proteinuria in human renal disease. *Lab. Invest.* 28:477–81
39. Blau, E. B., Michael, A. F. 1972. Rat glomerular glycoprotein composition and metabolism in aminonucleoside nephrosis. *Proc. Soc. Exp. Biol. Med.* 141:164–72
40. De Bats, A., Gordon, A. H., Rhodes, E. L. 1974. Variation in glomerular sialic acid content in diabetes and as the result of aging. *Clin. Sci. Mol. Med.* 47:93–95
41. Brenner, B. M., Baylis, C., Deen, W. M. 1976. Transport of molecules across renal glomerular capillaries. *Physiol. Rev.* 56:502–34

Ann. Rev. Med. 1977. 28:103–110
Copyright © 1977 by Annual Reviews Inc. All rights reserved

CYTOMEGALOVIRUS (CMV) IN THE COMPROMISED HOST(S)[1]

❖7227

Robert F. Betts, M.D.
Department of Medicine, University of Rochester School of Medicine and Dentistry, Rochester, New York 14642

James B. Hanshaw, M.D.
Department of Pediatrics, University of Massachusetts School of Medicine, Worcester, Massachusetts 01604

INTRODUCTION

An understanding of the role of cytomegalovirus (CMV) in the compromised host has increased with the understanding of the epidemiology of CMV in normal individuals. Studies using complement-fixing (CF) antibody tests have shown that there is a higher incidence of latent infection with CMV in women compared to men (1), an increasing incidence of latent infection with increasing age (2–4), and an increasing incidence related to sexual activity (2). Furthermore, this incidence varies in studies of different normal populations in different sections of the country (3, 4). However, serologic surveys have uniformly been carried out using the CF test, which is now known to be relatively insensitive. It can be negative in patients shedding virus (5), and in individuals with positive neutralizing or fluorescent-antibody (FA) activity (6). Better definition of the incidence of infection awaits studies with the FA technique that has recently been demonstrated to detect virtually all individuals with latent or active infection (6).

Two types of CMV infection occur in the compromised host. One type appears to be reactivation of infection acquired earlier in life, and the other is infection acquired at the time the immune mechanisms of the host are compromised. This review considers the differences in these two types of infections among compromised patients and examines why these different patterns of infection occur.

---

[1] Supported in part by a contract (NIH NIAMD 3-2218) from the Artificial Kidney and Chronic Uremia Program of the National Institute of Arthritis and Metabolic Diseases of the National Institutes of Health.

INFECTION IN IMMUNE DEFICIENCY

When the syndromes of combined immune deficiencies were first recognized, it became apparent that CMV was one of many agents capable of producing disease in this population (7). There are at least two reasons why CMV infection occurs in this setting: first, a significant percentage of pregnant women (10–20%) shed CMV in their cervix and their offspring acquire infection at birth (8); second, household contact is sufficient to allow transmission of infection among normal children (9), and therefore to children with immune deficiency. Although CMV transmitted to normal children either at birth or from their siblings is usually a benign or asymptomatic disease (8, 9), for the immune compromised child, dissemination and death may ensue. The relative infrequency of CMV disease in agammaglobulinemia suggests that cellular immunity or combined deficiency is of greater importance than humoral deficiency alone.

INFECTION IN NEOPLASIA

Leukemic children have antibody to CMV at a frequency similar to normal children of the same age. It has been shown that 7–25% of children with leukemia shed CMV (5, 10, 11). The higher frequency is noted when children are cultured repeatedly (5). Since most leukemic children who shed CMV are seropositive when first seen, and the first specimen obtained yields virus, most virus shedding is due to reactivation of latent infection acquired earlier in life. The explanation for shedding among individuals whose initial serologic test is negative is either that the CF test is insensitive (6) or that primary infection is acquired from other leukemic children or from blood products. Since there is strong epidemiologic evidence that CMV is occasionally transmitted by blood transfusion (12), probably both explanations hold, but the questions will not be resolved until a prospective study is carried out using the FA technique. It is possible that the slightly increased incidence of rash, pneumonitis, and fever noted in one study in infected compared to uninfected patients was provided by the children who had developed primary infection (5).

In adults, Duvall and co-workers recognized that CMV excretion in seropositive patients was very common in older individuals with solid tumor and lymphoproliferative disorders and that shedding of virus could not be linked to any recognizable disease process (13), presumably because shedding was related to reactivation of latent infection. Because the patients in this study were older (1–4), the frequency of antibody was higher than in the childhood leukemia group, but still did not approach the frequency noted in renal-transplant recipients (6, 14–16).

Although most adults with neoplasia and children with leukemia who shed virus are asymptomatic, some CMV infections result in generalized disease, usually diagnosed at autopsy (5, 13, 17) in conjunction with other morphologically defined infections. Most of these patients have relatively far-advanced neoplasia. The co-occurrence of CMV with other infectious agents at autopsy suggests that endogenous agents become disseminated because of a final deterioration in host defense rather than that primary infection occurs with multiple agents simultaneously. An intriguing but unproven alternative that would explain why CMV infection is superimposed so often on another morphologically obvious infectious disease is that in the severely ill patient, CMV produces further immunosuppression and allows emergence of the additional opportunistic organism, a phenomenon noted in other experimental viral infections in animals (18).

INFECTION IN RHEUMATIC DISEASE

It has now been clearly shown that intervention with cytotoxic drugs in patients with rheumatic disease may lead to CMV shedding (19). However, neither steroids alone nor the presence of a disease sufficiently severe to warrant intervention with steroids or cytotoxic drugs is sufficient to produce virus shedding by itself. Furthermore, not all patients who have antibody and who receive cytotoxic drugs shed virus; as has been noted in patients with neoplasia, those patients who shed virus are asymptomatic.

INFECTION IN PATIENTS UNDERGOING TRANSPLANTATION

The group of compromised hosts in which CMV has gained most attention is that of the recipients of transplanted organs. Infection has been recognized in recipients of cardiac (20), hepatic (21), bone-marrow (22), and renal allografts (6, 14–16, 23–25). In the former two instances, no systematic studies have been carried out to determine the frequency of infection and to determine whether virus shedding occurs only in pre-transplant seropositive patients. Many patients who die have morphologic evidence of CMV infection as well as evidence of infections with other agents. Deaths related to CMV infection appear to occur frequently in hepatic- and cardiac-transplant patients, which may be related to the higher dose of cytotoxic agents administered compared to other patients receiving cytotoxic drugs. Instead of simply shedding CMV, as do patients with neoplasia (13) or rheumatic disease (19), cardiac and hepatic patients develop dissemination of CMV.

The incidence of CMV infection in bone-marrow recipients may be higher than in leukemic children or in normal children. However, the bone-marrow recipients reported have had a frequency of latent infection prior to transplantation that was higher than in the studies of leukemic children. In bone-marrow transplants, the major thesis raised by Meyers and co-workers (22) is that CMV is etiologic in patients with interstitial pneumonia. Furthermore, it is implied that transmission of virus at transplantation may play a role, since patients with pneumonia received marrow from a seropositive donor more often than patients without pneumonia, although the authors only reported these observations in 22 of 85 transplanted patients. The data presented do not permit the conclusion as to whether primary infection occurred among allograft recipients. It should be noted, however, that pneumonia did occur in many patients with antibody prior to transplantation (22).

In renal-transplant recipients, a group that has been extensively studied, CMV has been demonstrated in 60–85% of recipients (6, 14–16, 23–25). This is the highest incidence of CMV infection in any group of compromised hosts. Since a number of shedders are asymptomatic, Craighead and co-workers were unable to define a typical clinical syndrome associated with infection (14). However, a number of other studies have reported fever, leukopenia, mononucleosis-like illness, pneumonia, and hepatitis due to CMV (15, 16, 23–26). Until recently the discrepancies in the observations were unresolved. Recently the answer to several of these questions was demonstrated simultaneously by two laboratories (6, 25). Infection results from reactivation of latent infection, or CMV is transmitted with the transplanted kidney. Seronegative individuals who receive a kidney from a seropositive donor undergo seroconversion and shed virus. When the FA test is used to designate serologic status, most but not all seronegative recipients who receive a kidney from a seropositive donor become infected, whereas seronegative recipients who receive a kidney from a seronegative donor do not develop infection after transplantation (6). Illness develops in approximately 90% of the individuals who acquire a primary infection with the transplanted kidney in the early months following transplantation. Symptoms may include fever, SGOT elevation, and pneumonia, and occur coincidentally with the immune response development (26) (see Figure 1). During that same period, illness is infrequent in patients who had antibody prior to transplantation in spite of the fact that most shed virus following surgery (6, 25, 26). Presumably, they are reactivating latent infection. Patients who have no CMV infection after transplantation are generally asymptomatic unless bacterial infection develops.

Thus these observations explain why the incidence of infection in transplant recipients is higher than in other compromised hosts. Not only does reactivation occur in the group of patients who have acquired infection throughout their life, but it occurs in approximately half of the patients who at the time of transplantation were free of CMV infection. Since kidneys transplanted from random donors would be derived half the time from seropositive individuals, seronegative recipients have a 50% chance of developing primary infection. Overall, one would predict a 75% incidence of virus shedding following transplantation, an incidence that has, in fact, been shown in several studies. Closer analysis has revealed that children who receive kidneys from their parents have a much higher incidence of primary infection than do recipients of cadaver or sibling kidneys. This is due to the higher frequency of latent infection in the older parent and low frequency in the young child, providing the ideal pairing for transmission of primary infection. Cadaver recipients have a low incidence of primary infection, since their donors are usually younger than they are and have a relatively low incidence of latent infection. Sibling pairs are of nearly identical age, and have similar environmental exposures and similar antibody titers.

Finally, two interesting observations have been made in allograft recipients. In marrow recipients, graft-versus-host reaction occurs more frequently in patients with pneumonitis than in those without it (22). Since CMV has been implicated in pneumonitis, this would raise the question of the role of the virus in that reaction. In renal transplants, rejection has been associated with the development of CMV infection (16, 24, 26). Although it is possible that symptoms of rejection cause the physician to alter the immunosuppressive program and hence unmask the virus

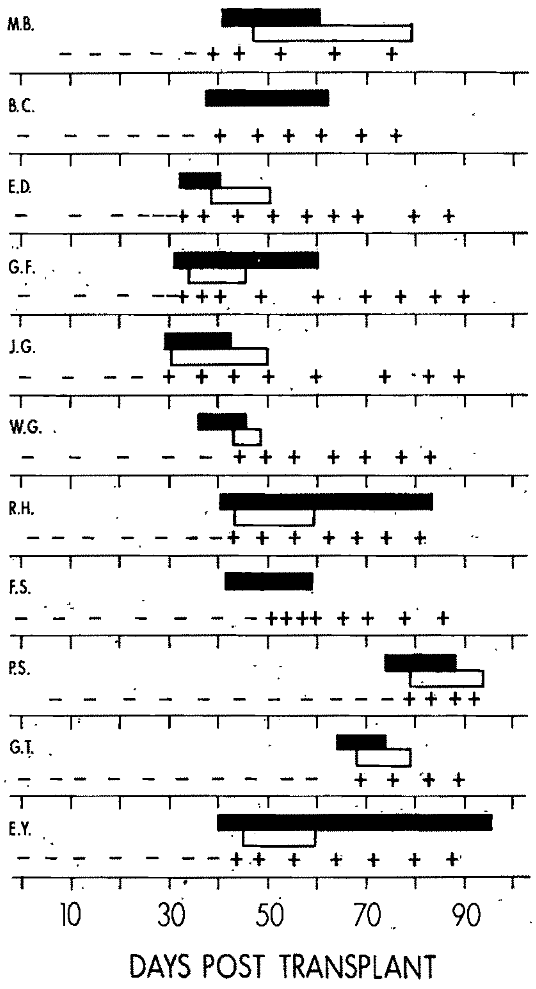


Figure 1 Development of fever ■ and SGOT rise ▭ in relation to conversion from negative (–) to positive (+) antibody to CMV in eleven patients with primary CMV infection.

infection, the bulk of the evidence does not support that possibility. Rejection and infection occur before any major change in immunosuppression in many instances. A possible explanation for these observations is that viral infection is producing a change in the surface antigens, perhaps interacting with the HL-A antigens. This leads, in the case of bone-marrow transplants, to a more fulminant rejection of the host by the graft, or, in the case of the renal transplant, to a more fulminant rejection of the kidney. Although no data pertinent to this question are available in man, Schrader and co-workers have proposed that the H-2 locus in animals may serve to incorporate foreign antigens into its structure and provide a site for attack and elimination by cytotoxic lymphocytes (27). Although the relationship of graft-versus-host or renal rejection in CMV infection remains speculative, it provides an important subject for future investigation.

FACTORS CONTRIBUTING TO VIRUS SHEDDING

An important area of host-virus interaction that remains entirely unclear relates to the factor(s) that lead to reactivation of virus and then to persistence of shedding once reactivation has occurred. At the present time there is only indirect evidence to suggest that CMV reactivation occurs spontaneously in man. The presence of virus in cervical secretions of normal women supports a concept of reactivation, but since cervical shedding is high in young women and low in older women, virus shedding may represent continued excretion following a primary infection (8). Quite clearly, however, reactivation follows the use of cytotoxic agents and steroids in patients, and alkylating agents appear to be more important than steroids (19). Shedding, once it develops, often persists for months or even years. By contrast, reactivation of herpes simplex shedding is of short duration. Typically, the renal-transplant recipient or the child with leukemia sheds herpes simplex in oral secretions shortly after transplantation or treatment. Herpes will then disappear, and CMV will appear and then persist for months without the reappearance of herpes simplex. On the other hand, in normal subjects with antibody to herpes and CMV, herpes simplex shedding, but not CMV shedding, occurs intermittently (28). One explanation for this difference might relate to the fact that cellular mechanisms by themselves are responsible for control of CMV, and antibody plays no role in protection but acts only as a marker of infection. In contrast, both cellular and humoral immunity appear to contribute to control of herpes simplex infection. Since one leg of the system remains intact in the individual placed on cytotoxic agents, herpes simplex virus shedding is only intermittent, whereas CMV shedding is prolonged. Further study of this interesting contrast could provide important information concerning host response to viral infection.

CONCLUSIONS

Cytomegalovirus infection in compromised hosts occurs most frequently as asymptomatic shedding of a reactivated virus. Reactivation occurs when cellular immunity is compromised owing to either immune-deficiency disease or iatrogenic intervention. When intervention is modest, most antibody-positive patients do not shed virus (19); when the group is treated with a higher mean dose of cytotoxic agents, almost all antibody-positive patients shed virus (6). In the instance where the chemotherapeutic program includes massive doses of these agents, as in heart or marrow transplantation, or when dysfunction of all cellular activity is extreme, as in the case of severe combined immune deficiency and far-advanced neoplasia, dissemination of infection with multiple-organ involvement ensues (17, 20–22).

Primary infection may occur in some patients with neoplasia or in marrow-transplant recipients from either blood transfusions or from marrow, although data are insufficient to be certain (22). It has been conclusively shown, however, that seronegative renal-allograft recipients usually develop primary infection when they acquire a kidney from a seropositive donor. Clinical illness related to virus infection in the first few months after transplantation occurs in almost all patients with primary infection and seldom in those with reactivation infection (6, 25, 26). Three very intriguing observations have been made that should suggest avenues for future research: (*a*) Since illness occurs more closely related to development of antibody than it does to onset of virus shedding, this suggests that host response plays an important role in the disease that is obviously caused by a virus; a definition of the mechanism of this interaction could lead to an understanding of host-induced disease processes in man; (*b*) Graft-versus-host disease in marrow transplantation; and (*c*) allograft rejection in renal-graft recipients correlate with development of CMV infection (16, 22, 24, 26). Taken together, these observations are consistent with the hypothesis that virus antigens interrelate with most antigens in such a way as to lead to immune response to both antigens. Exploration of this clinical observation could lead to a greater understanding of the function of the HL-A system in man.

Literature Cited

1. Luby, J. P., Shasby, D. M. 1972. A sex difference in the prevalence of antibodies to cytomegalovirus. *J. Am. Med. Assoc.* 222:1290–91
2. Davis, L. E., Stewart, J. A., Garvin, S. 1975. Cytomegalovirus infection: A seroepidemiologic comparison of nuns and women from a venereal disease clinic. *Am. J. Epidemiol.* 102:327–30
3. Mirkovic, R., Werch, J., South, M. A., Benyesh-Melnick, M. 1971. Incidence of cytomegaloviremia in blood bank donors and infants with congenital cytomegalic inclusion disease. *Infect. Immun.* 3:45–50
4. Hanshaw, J. B. 1968. Cytomegaloviruses. *Monogr. Virol.* 3:1–23
5. Henson, D., Siegel, S. E., Fuccillo, D. A., Mathew, E., Levine, A. S. 1972. Cytomegalovirus infection in childhood leukemia. *J. Infect. Dis.* 126:469–81
6. Betts, R. F., Freeman, R. B., Douglas, R. G. Jr., Talley, T. E., Rundell, B. 1975. Transmission of cytomegalovirus infection with renal allograft. *Kidney Int.* 8:837–94
7. Hoyer, J. R.,Cooper, M. D., Gabrielsen, A. E., Good, R. A. 1968. Lymphopenic forms of congenital immunologic deficiency: Clinical and pathologic patterns. *Birth Defects Orig. Artic. Ser.* 4:91–103
8. Stagno, S., Reynolds, D. W., Tsiantos, A., Fuccillo, D. A., Long, W., Alford, C. A. 1975. Comparative serial virologic and serologic studies of symptomatic and subclinical congenitally and naturally acquired cytomegalovirus infection. *J. Infect. Dis.* 132:568–77
9. Hanshaw, J. B., Betts, R. F., Simon, G., Boynton, R. C. 1965. Acquired cytomegalovirus infection. *N. Engl. J. Med.* 272:602–9
10. Benyesh-Melnick, M., Dessy, S. I., Fernbach, D. J. 1964. Cytomegaloviruria in children with acute leukemia and in other children. *Proc. Soc. Exp. Biol. Med.* 117:624–30
11. Renker, C. T., McGraw, J. P. 1967. Cytomegalic inclusion disease in childhood leukemia. *Cancer* 20:36–39
12. Kääriäinen, L., Klemola, E., Paloheimo, J. 1966. Rise of cytomegalovirus antibodies in an infectious mononucleosis-like syndrome after transfusion *Br. Med. J.* 1:1290–92
13. Duvall, C. P., Casazza, A. R., Grimley, P. M., Carbone, P. P., Rowe, W. P. 1966. Recovery of cytomegalovirus from adults with neoplastic disease. *Ann. Intern. Med.* 64:531–41
14. Craighead, J. E., Hanshaw, J. B., Carpenter, C. B. 1967. Cytomegalovirus in-

fection after renal transplantation. *J. Am. Med. Assoc.* 201:725–28

15. Spencer, E. S. 1974. Clinical aspects of cytomegalovirus infection in kidney graft recipients. *Scand. J. Infect. Dis.* 6:315–23
16. Lopez, C., Simmons, R. L., Mauer, S. M., Najarian, J. S., Good, R. A., Gentry, S. 1974. Association of renal allograft rejection with viral infections. *Am. J. Med.* 56:280–89
17. Macasaet, F. F., Holley, K. E., Smith, T. F., Keys, T. F. 1975. Cytomegalovirus studies of autopsy tissue. *Am. J. Clin. Pathol.* 63:859–65
18. Notkins, A. L., Mergenhagen, S. E., Howard, R. J. 1970. Effects of virus infections on the function of the immune system. *Ann. Rev. Microbiol.* 24:525–38
19. Dowling, J. N., Saslow, A. R., Armstrong, J. A., Ho, M. 1976. Cytomegalovirus infection in patients receiving immunosuppressive therapy for rheumatic disorders. *J. Infect. Dis.* 133:399–408
20. Stenson, E. B., Bieber, C. P., Griepp, R. B., Clark, D. A., Shumway, N. E., Remington, J. S. 1971. Infectious complications after cardiac transplantation in man. *Ann. Intern. Med.* 74:22–36
21. Fulginiti, V. A., Scribner, R., Groth, C. G., Putnam, C. W., Brettschneider, L., Gilbert, S., Porter, K. A., Starzl, T. E. 1968. Infections in recipients of liver homografts. *N. Engl. J. Med.* 279:619–26
22. Meyers, J. D., Spencer, H. C., Watts, J. C., Gregg, M. B., Stewart, J. A., Troupin, R. H., Thomas, E. D. 1975. Cytomegalovirus pneumonia after human marrow transplantation. *Ann. Intern. Med.* 82:181–88
23. Luby, J. P., Burnett, W., Hull, A. R., Ware, A. J., Shorey, J. W., Peters, P. C. 1974. Relationship between cytomegalovirus and hepatic function abnormalities in the period after renal transplant. *J. Infect. Dis.* 129:511–18
24. Fiala, M., Payne, J. E., Berne, T. V., Moore, T. C., Henle, W., Montgomerie, J. Z., Chaterjee, S. N., Guze, L. B. 1975. Epidemiology of cytomegalovirus infection after transplantation and immunosuppression. *J. Infect. Dis.* 132:421–33
25. Ho, M., Suwansirikul, S., Dowling, J. N., Youngblood, L. A. Armstrong, J. A. 1975. The transplanted kidney as a source of cytomegalovirus infection. *N. Engl. J. Med.* 293:1109–12
26. Betts, R. F., Freeman, R. B., Douglas, R. G., Pabico, R. C. 1975. Epidemiology and significance of cytomegalovirus infection in hemodialysis and transplant patients. *Clin. Res.* 23:355A
27. Schrader, J. W., Cunningham, B. A., Edelman, G. M. 1975. Functional interactions of viral on histocompatibility antigens at tumor cell surfaces. *Proc. Natl. Acad. Sci.* 72:5066–70
28. Douglas, R. G. Jr., Couch, R. B. 1970. A prospective study of chronic herpes simplex virus infection and recurrent herpes labialis in humans. *J. Immunol.* 104:289–95

Ann. Rev. Med. 1977. 28:111–17
Copyright © 1977 by Annual Reviews Inc. All rights reserved

DIAGNOSIS OF HYPERPARATHYROIDISM

❖7228

Henry G. Bone, III, M.D., William H. Snyder, III, M.D. and Charles Y. C. Pak, M.D.
Departments of Internal Medicine and Surgery, The University of Texas
Southwestern Medical School, Dallas, Texas 75235

Hyperparathyroidism is a disorder of excessive secretion of parathyroid hormone (PTH) by one or more glands that have undergone abnormal cellular proliferation, either as a primary event or in response to other physiologic abnormalities. We discuss the application of both long-established and recently developed methods to the diagnosis of hyperparathyroidism, and pay particular attention to the primary disorder.

CLINICAL PRESENTATION

The increasing use of multichannel automated testing, including determination of the serum calcium, the elevation of which is the hallmark of primary hyperparathyroidism, has resulted in the screening of a large portion of the population. It is now recognized that primary hyperparathyroidism is more common than was formerly appreciated. It is currently estimated that the prevalence of the disease is approximately one case per thousand population (1). As cases are discovered earlier in their course, the typical clinical presentation has become more subtle, and obvious osseous and renal disease are now seen in a minority of cases. The shifting spectrum of the clinical presentation of primary hyperparathyroidism is illustrated by a comparison of the findings in two series, one representing a long experience with the disease prior to 1966 (2) and the other, the recent experience of the authors (Table 1).

Special consideration must be given to the diagnosis of hyperparathyroidism in the context of several disorders with which it is often associated. Nephrolithiasis is commonly associated with hyperparathyroidism. In the authors' recent experience, parathyroid disease is found in 18% of patients with renal stones and these patients constitute 28% of patients with hyperparathyroidism (Table 1.) The severe bone disease of hyperparathyroidism occurs mainly in post-menopausal women (3), and

Table 1 Clinical presentation of patients with primary hyperparathyroidism (Percentage of total cases)

Presenting feature	Cope ($n = 342$)	Authors[a] ($n = 75$)
Overt bone disease	23	13
Nephrolithiasis	57	28
Peptic ulcer	8	8
Asymptomatic	1	20
Hypertension	2	47
Other	9	—

[a] In the authors' series, some patients had more than one presenting complaint. Many of our cases were discovered by screening a clinic population in which hypertension is the most common diagnosis. If hypertension is considered to be largely unrelated to hyperparathyroidism, a somewhat larger number of our patients should be regarded as asymptomatic.

should be considered in the evaluation of osteoporosis. Dyspepsia and peptic-ulcer disease may result from hypercalcemia. Hyperparathyroidism is an important feature of the syndromes of multiple endocrine neoplasia (MEN), particularly type I, in which it is associated with other endocrine adenomas, including the Zollinger-Ellison syndrome.

PATHOPHYSIOLOGY

The term *primary hyperparathyroidism* is used to denote exaggerated parathyroid function that appears to arise independently of any physiologic stimulus. This may be due to an adenoma (or possibly adenomas in cases of multiple endocrine adenomatosis), to hyperplasia or, uncommonly, to carcinoma. Although it is not clear why generalized hyperplasia should occur as a primary abnormality, in many series the diagnosis of primary hyperplasia has been made in approximately 10–20% of cases (4, 5). Interestingly, in a recent study biopsies were taken from every parathyroid gland identified, and microscopic evidence of hyperplasia was found in about one half of cases (6).

The states in which parathyroid hyperfunction occurs in response to hypocalcemic influences are termed secondary hyperparathyroidism. In certain forms of rickets and osteomalacia, secondary hyperparathyroidism plays a major role (7). The most important cause of secondary hyperparathyroidism is renal failure, in which PTH secretion is stimulated by effects of hyperphosphatemia and the decreased intestinal absorption of calcium (Ca) associated with impaired renal hydroxylation of vitamin D (8). Renal hypercalciuria also stimulates secretion of PTH (9). An important feature in some cases of secondary hyperparathyroidism is the development of autonomy (sometimes called "tertiary hyperparathyroidism"), in which hypersecretion of PTH is out of proportion to the stimulus or persists long after the removal of the stimulus. The usual setting for this problem is renal failure, in which PTH secretion may become sufficient to produce hypercalcemia. This is most often manifested after transplantation.

In MEN type II, the medullary thyroid carcinoma/pheochromocytoma syndrome, the hyperparathyroidism may represent a hyperplastic response to the hypocalcemic effects of calcitonin.

Among the causes of hypercalcemia in patients with various malignancies is the elaboration by the tumor of PTH or substances with similar biologic effects. Such PTH-like substances may be associated with relatively low circulating levels of immunoreactive PTH (iPTH) (10), suggesting that in some instances there are significant structural differences between the native hormone and that produced by the tumor. The possibility of such a malignancy should be considered in patients with biochemical findings compatible with hyperparathyroidism.

DIAGNOSTIC TESTS

The physiologic abnormalities produced by the presence of excessive amounts of PTH are directly reflected by several abnormal serum chemical determinations, which are widely available and are often included in multichannel screening panels. Of these, by far the most important is the serum calcium. Keating (11) emphasized that abnormalities of the serum calcium concentration usually imply disturbed parathyroid function. He demonstrated that when careful attention was paid to the serum calcium level, an elevated mean value constituted a generally reliable indication of hyperparathyroidism, if other etiologies were excluded (specifically, malignant disease, sarcoidosis, hyperthyroidism, hypervitaminosis D, milk-alkalai syndrome, Paget's disease of bone with immobilization, and acute osteoporosis of disuse). It should be recognized that normal serum Ca levels decrease with age, especially in men (12). Another effect of PTH is the reduction of the serum P level, largely by enhancement of urinary excretion, but in many patients this effect is not sufficient to produce overt hypophosphatemia. PTH increases renal bicarbonate excretion and tends to produce a hyperchloremic acidosis, which is usually mild. Some authors have attempted to magnify the diagnostic utility of the hyperchloremic and hypophosphatemic effects of PTH by employing the chloride (Cl)-to-phosphorus (P) ratio as a diagnostic parameter (13–15). However, in most of the cases examined the diagnoses were reasonably clear-cut on other grounds, so the utility of the Cl:P ratio in ambiguous cases is still unclear. In our experience there is an appreciable overlap between the Cl:P ratios of patients with hyperparathyroidism and normal controls. Measurements of the phosphaturic effects of PTH, expressed in terms of tubular reabsorption of phosphorus or as renal clearance of phosphorus, have also been employed in the diagnosis of hyperparathyroidism. However, the results of these tests are often not conclusive, since variations in P intake, serum P levels, and renal function limit the reliability of these studies (16).

Recent contributions to the diagnostic armamentarium include the determinations of urinary cyclic adenosine monophosphate (cAMP), and of serum iPTH. The renal effects of PTH are significantly reflected by the urinary cAMP (17, 18). This determination is usually made on a specimen of urine collected over a 24-hr period, and thus provides a value that is in effect integrated over time. A recent report indicates that specific determination of nephrogenous cAMP enhances the diagnostic power of the urinary cAMP determination (19). There are several factors that may limit the significance of the urinary cAMP. Increased levels may be found in diabetes mellitus, thyrotoxicosis, acromegaly, states of increased antidiuretic hormone activity, and conditions marked by adrenergic excess. Drugs that impair renal tubular reabsorption of Ca also increase cAMP excretion. Depressed urinary cAMP levels are often seen in patients with renal insufficiency even in the face of hyperparathyroidism. Conditions of diet and drug intake should be carefully controlled in the evaluation of patients by means of urinary cAMP. The main virtues of the cAMP determination are the relative simplicity and reliability of the assay techniques.

The determination of the serum iPTH level potentially provides the most direct measure of parathyroid function. However, several problems have limited the availability and reliability of this measurement. Although structurally distinct from human PTH, bovine PTH is generally used as the tracer in radioimmunoassays; consequently, the systems are heterologous. Intact hormone and various fragments of PTH from several species have been used as antigens to develop antisera in several animal species. These antisera vary widely as to sensitivity and specificity. Antisera to the longer-lived C-terminal fragments may give information of a relatively time-integrated character, while N-terminal systems are more sensitive to fluctuations in secretory rates (20). For these reasons, normal values must be established in each laboratory. Interpretation of iPTH levels also requires consideration of the serum calcium level; it is important to recognize that iPTH levels that might be considered normal in a normocalcemic patient may be inappropriately high in a patient with hypercalcemia (20). With most assay techniques, detectable levels of iPTH, in the presence of hypercalcemia, suggest hyperparathyroidism. For this reason it is important that the assay system used be sensitive enough to measure iPTH in the sera of a fair proportion of normal subjects. Otherwise, iPTH may be unmeasurable in a significant number of patients with hyperparathyroidism. Elevation of the iPTH level in normocalcemic patients does not establish the diagnosis of primary hyperparathyroidism, as elevated levels are found in secondary hyperparathyroidism.

Standard roentgenologic evaluation may be of some diagnostic value, and is of use in assessment of the severity and duration of the disease. The characteristic signs, especially subperiosteal bone resorption, are often seen in advanced cases, but are only seen with special techniques in most patients. Angiographic techniques may be combined with the radioimmunoassay determination of PTH levels in specimens obtained by selective catheterization of neck veins in order to localize parathyroid tumors (21). Such an approach is technically very demanding and, if undertaken sporadically, may fail to provide helpful information.

Bone densitometry by special roentgenographic methods or by ^{125}I-photon absorption is a useful tool in assessing the degree of loss of bone mineral. Among hyperparathyroid patients the most striking abnormalities are seen in postmenopausal white women (3).

SELECTION OF PATIENTS FOR SURGERY

Increasingly attentive and detailed observations of asymptomatic patients with primary hyperparathyroidism have revealed that many of them have slightly impaired renal function or significant physiologic derangements, such as negative Ca balance, hypercalciuria, low bone density, or intestinal hyperabsorption of Ca, which are corrected by parathyroidectomy (22). We suspect that without intervention, many, if not all, such patients would undergo progression of their disease. A small number of patients have no symptoms or discernible physiologic abnormalities beyond mild hypercalcemia. Longitudinal study of a group of such patients over several years will be necessary in order to characterize the course of their disease. Except for these individuals and those with major contraindications, it is our practice to recommend parathyroidectomy. It is essential that regular comphrehensive reevaluation be carried out for cases in which operation is withheld.

Histopathological evaluation of diseased parathyroid tissue is often problematical. The classical division between adenoma and hyperplasia is no longer a simple one, and the diagnosis that seems most consistent with the microscopical findings may not fit the clinical circumstances very well. The clinical setting, particularly the family history, may be crucial to the correct diagnosis, as may the observations of the experienced surgeon. The problem of the "tissue diagnosis" is reviewed elsewhere (6, 23–25). We find that the operational terms "single-gland involvement" and "multiple-gland involvement" as used by Paloyan & Lawrence (23) are quite useful.

INTRAOPERATIVE EVALUATION

Intraoperative evaluation is directed toward discrimination between single-gland and multiple-gland disease. The latter category includes patients with multiple endocrine neoplasia, familial hyperparathyroidism, and renal insufficiency, as well as those patients in whom multiglandular disease is diagnosed intraoperatively. This distinction is significant because it determines the choice between resection of a single gland and subtotal resection of the mass of parathyroid tissue (leaving approximately 50 mg). It is our practice to perform the latter operation whenever there appears to be actual or potential multiglandular disease.

Because of the limitations of histopathological evaluation, especially by frozen section, the assessment of the parathyroid glands depends in a major way upon the surgeon's gross evaluation. All glands should, if possible, be identified and evaluated prior to any resection. Abnormalities of appearance and consistency may be helpful to the experienced parathyroid surgeon, but excessive glandular size probably provides the most reliable indication of pathology. Since their weight cannot be determined in situ, the glands must be measured prior to resection. A gland larger than 6 mm in its greatest dimension usually weighs more than 50 mg and should therefore be considered abnormal. All resected tissue should be weighed. If two or more glands are enlarged, we consider the disease to be multiglandular.

The finding of a dense, tough, adherent parathyroid gland suggests the diagnosis of carcinoma, which cannot always be confirmed histologically if such findings as invasion are not seen.

SUMMARY

States of hypersecretion of PTH may occur primarily, or in response to other physiologic abnormalities. Primary hyperparathyroidism must be considered in the differential diagnosis of hypercalcemia, nephrolithiasis, metabolic bone disease, and pancreatitis and peptic-ulcer disease. The clinical manifestations of this disease have become more subtle with improved detection. The serum calcium level is almost always elevated, and when it is accompanied by relatively high serum PTH levels or increased urinary cAMP excretion, the diagnosis is usually secure. Findings of hypophosphatemia, decreased renal tubular reabsorption of phosphorus, hypercalciuria, and characteristic roentgenographic changes support the diagnosis of hyperparathyroidism, but are not prerequisites for that diagnosis. Most cases will come to operation, and experienced intraoperative assessment is necessary for the correct distinction between multiglandular disease and that involving only a single gland. We expect that a clearer understanding of the histopathologic features of these diseases, and improvement in the methods for measurement of PTH will be the main areas of advancement in the diagnosis of hyperparathyroidism in the next few years.

ACKNOWLEDGMENT

The authors wish to acknowledge the support of USPHS grants RO1-AM-16061, MO1-RR-00633, and HL-11662.

Literature Cited

1. Boonstra, C. E., Jackson, C. E. 1971. Serum Ca survey for hyperparathyroidism. *Am. J. Clin. Pathol.* 55:523–26
2. Cope, O. 1966. The story of hyperparathyroidism at the Massachusetts General Hospital. *N. Engl. J. Med.* 274:1174–82
3. Pak, C. Y. C., Kaplan, R., Notz, C., Stewart, A., Bone, H., Browne, R. 1975. Photon absorptiometric analysis of bone density in primary hyperparathyroidism. *Lancet* 2:7–8
4. Block, M. E., Frame, B., Jackson, C. E., Horn, R. C. 1974. The extent of operation for primary hyperparathyroidism. *Arch. Surg.* 109:789–801
5. Rasmussen, H. 1974. In *Textbook of Endocrinology*, ed. R. H. Williams, pp. 660–773. Philadelphia: Saunders. 1138 pp.
6. Esselstyn, C. B., Levin, H. S., Eversman, J. J., Schumacher, O. P., Skillern, P. G. 1974. Reappraisal of parathyroid pathology in hyperparathyroidism. *Surg. Clin. North Am.* 54:443–47
7. Scriver, C. R. 1974. Rickets and the pathogenesis of impaired tubular transport of phosphate and other solutes. *Am. J. Med.* 57:43–49
8. Bricker, N. S., Slatopolsky, E., Reiss, E., Avioli, L. V. 1969. Calcium, phosphorus, and bone in renal disease and transplantation. *Arch. Intern. Med.* 123:543–53
9. Coe, F. L., Canterbury, J. J., Firpo. J. J., Reiss, E. 1973. Evidence for secondary hyperparathyroidism in idiopathic hypercalciuria. *J. Clin. Invest.* 52: 134–42
10. Benson, R. C., Jr., Riggs, B. L., Pickard, B. M., Arnaud, C. D. 1974. Radioimmunoassay of parathyroid hormone in hypercalcemic patients with malignant disease. *Am. J. Med.* 56:821–26
11. Keating, R. F. 1961. Diagnosis of primary hyperparathyroidism. *J. Am. Med. Assoc.* 178:547–55

12. Keating, R. F., Jones, J. D., Elveback, L. R., Randall, R. V. 1969. The relation of age and sex to distribution of values in healthy adults of serum calcium, inorganic phosphorus, magnesium, alkaline phosphatase, total proteins, albumin and blood urea. *J. Lab. Clin. Med.* 73:825–34
13. Palmer, F. J., Nelson, J. C., Bacchus, H. 1974. The chloride-phosphate ratio in hypercalcemia. *Ann. Intern. Med.* 80: 200–4
14. Reeves, C. D., Palmer, F., Bacchus, H., Longerbeam, J. K. 1975. Differential diagnosis of hypercalcemia by the chloride/phosphate ratio. *Am. J. Surg.* 130:166–71
15. Drago, J. R., Rohner, T. J. Jr., Sanford, E. J., Santen, R. E. 1976. Diagnosis of hyperparathyroidism. *Urology* 7:4–6
16. Massry, S. G., Freidler, R. M., Coburn, J. W. 1973. Excretion of phosphate and calcium. *Arch. Intern. Med.* 131:828–59
17. Murad, F., Pak, C. Y. C. 1972. Urinary excretion of adenosine 3', 5'-monophosphate and guanosine 3', 5'-monophosphate. *N. Engl. J. Med.* 286:1382–87
18. Kaminsky, N. I., Broadus, A. E., Hardman, J. G., Jones, D. J. Jr., Ball, U. H., Sutherland, E. W., Liddle, G. W. 1970. Effects of parathyroid hormone on plasma and urinary adenosine 3', 5'-monophosphate in man. *J. Clin. Invest.* 49:2387–95
19. Broadus, A. E., Mahaffey, J. E., Neer, R. M., Bartter, F. C. 1976. Nephrogenous cyclic AMP as a parathyroid function test. *Clin. Res.* 24:456A
20. Arnaud, C. D., Goldsmith, R. S., Bordier, P. J., Sizemore, G. W. 1974. Influence of immunoheterogeneity of circulating parathyroid hormone on results of radioimmunoassays of serum in man. *Am. J. Med.* 56:785–93
21. Bilezikian, J. P., Doppman, J. L., Shimkin, P. M., Powell, D., Wells, S. A., Heath, D. A., Ketcham, A. S., Mouchik, J., Mallette, L. E., Potts, J. T. Jr., Aurbach, G. D. 1973. Preoperative localization of abnormal parathyroid tissue: cumulative experience with venous sampling and arteriography. *Am. J. Med.* 55:505–14
22. Kaplan, R. A., Snyder, W. H., Stewart, A., Pak, C. Y. C. 1976. Metabolic effects of parathyroidectomy in asymptomatic primary hyperparathyroidism. *J. Clin. Endocrinol. Metab.* 42:415–26
23. Paloyan, E., Lawrence, A. M. 1974. Laboratory diagnosis of parathyroid tumors. *Ann. Clin. Lab. Sci.* 4:241–49
24. Strauss, F. H. II, Paloyan, E. 1969. The pathology of hyperparathyroidism. *Surg. Clin. North Am.* 49:27–42
25. Williams, E. D. 1974. Pathology of the parathyroid glands. *Clin. Endocrinol. Metab.* 3:285–303

Ann. Rev. Med. 1977. 28:119–30
Copyright © 1977 by Annual Reviews Inc. All rights reserved

❖7229

THE ROLE OF GLUCAGON IN THE ENDOGENOUS HYPERGLYCEMIA OF DIABETES MELLITUS

Roger H. Unger, M.D.
Veterans Administration Hospital and Department of Internal Medicine,
University of Texas, Southwestern Medical School, Dallas, Texas 75235

Lelio Orci, M.D.
Institute of Histology and Embryology, Geneva, Switzerland

INTRODUCTION

Recent advances in our understanding of the glucagon-secreting pancreatic α cell as a vital component of the functional machinery of the primary fuel-regulating organ, the islets of Langerhans, have made possible new insights into its physiology and pathophysiology. It has recently been proposed that glucagon, or closely related polypeptides with biologic actions similar to those of glucagon, plays an essential role in the development of the full diabetic syndrome, specifically, endogenous hyperglycemia (1, 2) and hyperketonemia (2, 3). If this is true, a safe and practical means of glucagon suppression might have therapeutic potential in improving the currently unsatisfactory methods of controlling diabetic hyperglycemia. The purpose of this review is to consider briefly the evidence that assigns to glucagon a vital role in the maintenance of euglycemia in nondiabetics and a deleterious role in certain of the metabolic abnormalities of diabetes mellitus.

NORMAL STRUCTURE AND FUNCTION OF THE GLUCAGON-SECRETING α CELLS OF THE ISLETS OF LANGERHANS

To comprehend fully the character of the disturbance of α-cell function in diabetes, one must understand its normal physiological role as a part of an organ that manages fuel distribution, the islets of Langerhans. The complex task of directing the flow of important nutrients to the different tissues in accordance with their varying needs for and the availabilities of the nutrients is accomplished with remarkable facility. An organ of fuel distribution must be able (*a*) to sense the availability of fuels and the needs of the various tissues, and (*b*) to respond bidirectionally either to promote the storage of fuels in abundance or to promote the retrieval of fuels when they are needed. As depicted in Figure 1, this is accomplished by varying the relative amounts of two biologically antagonistic secretory products of the islets of Langerhans: insulin, the hormone of fuel storage, and glucagon, the hormone of fuel retrieval. Thus the normally functioning alpha-beta–cell couple, in time of abundance of exogenous fuels, releases into the circulation a secretion mixture high in insulin relative to glucagon; this promotes the biosynthesis of glycogen, fat, and protein. Conversely, in time of need for endogenous fuel production, the normal islets produce a mixture rich in glucagon relative to insulin; this favors an increase in hepatic glucose production from glycogen and from gluconeogenic precursors and increased production of free fatty acids, glycerol, and ketones.

As Cahill has emphasized (4), the regulation of glucose flux in normal mammals differs to a striking degree from that of other fuels. Whereas the concentration of most fuels changes markedly during a change in their flux rate, in the case of glucose, flux rates may change drastically, but its extracellular concentration remains within the relatively narrow range that is regarded clinically as "normal." The purpose of avoiding hypoglycemia seems obvious—glucose, an obligatory fuel for the brain,

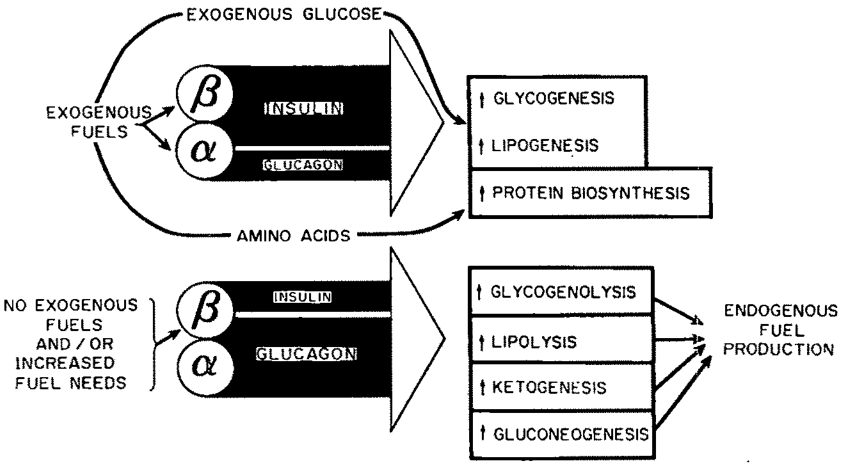


Figure 1 By secreting a high-insulin–low-glucagon mixture in time of nutrient abundance, the α-cell–β-cell unit promotes anabolic processes such as glycogenesis, lipogenesis, and protein biosynthesis. During need for endogenous fuel production, as in famine and fight and flight situations, a low-insulin–high-glucagon secretion mixture promotes retrieval of fuels from the macromolecules by catabolic processes such as glycogenolysis, gluconeogenesis, and lipolysis. Reprinted from *Diabetes* (5).

must be maintained above a certain critical level if the central nervous system is to function properly. However, the purpose of avoiding hyperglycemia is less apparent, but could be related to the possible role of hyperglycemia in the pathogenesis of diabetic microangiopathy. Whatever the reasons, regulation of glucose homeostasis in this manner requires a push-pull system that can control independently the rates of glucose flux into and out of the extracellular space, as depicted in Figure 2A (5). Glucagon, through its effects on hepatic glucose production, controls endogenous

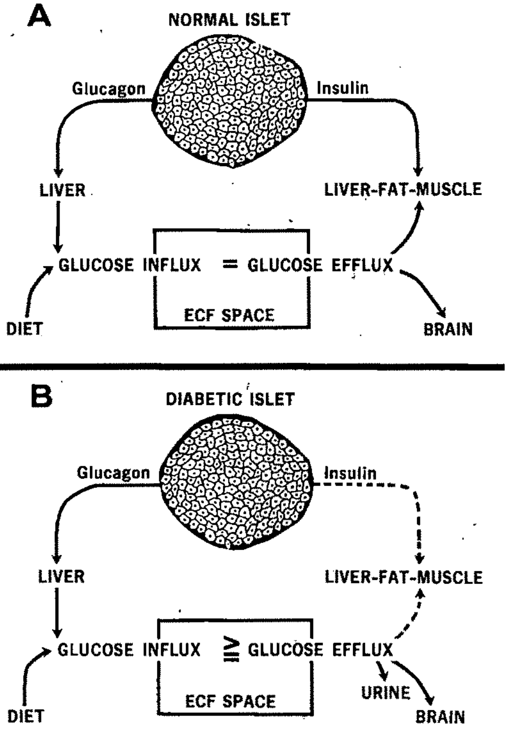


Figure 2 Models for the bihormonal control of extracellular fluid (ECF) glucose concentration by the normal islet (A) and the diabetic islet (B):

A. The glucose-sensing capacity of the normal islet maintains an insulin-glucagon mixture designed to restrict the range of extracellular-fluid glucose concentration irrespective of glucose flux rates. An increase in glucose influx is instantly met by an increase in insulin-induced glucose efflux so as to restore normoglycemia. Similarly, an increase in glucose efflux, resulting from exercise or protein ingestion, is met by a glucagon-mediated increase in glucose influx so as to prevent hypoglycemia.

B. The islet of Langerhans, in severe diabetes, has lost its glucose-sensing capacity and/or its ability to respond to glycemic change. Increases in glucose influx, whether from the diet or via glucagon-mediated enhancement of hepatic glucose production, cannot be matched by insulin-mediated increments in glucose efflux.

glucose influx and insulin controls glucose efflux into insulin-responsive tissues, liver, muscle, and fat. By varying the insulin-glucagon secretion mixture, the islets can maintain constant equality of glucose influx and efflux and thereby limit change in extracellular glucose concentration—even though glucose turnover may be changing widely. For example, during a high-glucose meal, 50 g of glucose may pour into the extracellular space within an hour, but glucose concentration will rise only transiently because an outpouring of insulin rapidly raises the rate of glucose efflux above that of glucose influx. When the glucose concentration is restored to basal levels, equality of influx and efflux is reestablished promptly. Similarly, when glucose efflux is increased, as during exercise or as the result of aminogenic insulin secretion from a protein meal, an appropriate outpouring of glucagon augments hepatic glucose production and prevents hypoglycemia. This changing of the insulin-glucagon mixture so as to maintain a balance of glucose production and glucose utilization must require a coordinated relationship of the glucagon-producing α cells and the insulin-secreting β cells under the guidance of a perceptive glucose sensor. This sensor must respond not only to change in glucose concentration, but must vary secretion responses to other nonglucose influences in accordance with glucose needs (i.e. the insulin and glucagon response to a protein meal is entirely different following a period of carbohydrate restriction than during a period of carbohydrate abundance) (6).

The anatomical equipment required to accomplish this functional feat is contained in microorgans scattered throughout the pancreas, the islets of Langerhans. Although the organization of the various cell types differs from species to species, in man the characteristic arrangement depicted in Figure 3*a* is well established (7, 8). It consists of an outer mantle of glucagon-containing α cells surrounding a homogeneous central mass of insulin-containing β cells comprising more than 60% of the islet cells. Intervening between the thin α-cell rim and the β-cell core are the D cells, recently identified (9–11) as the cells that contain the somatostatin or

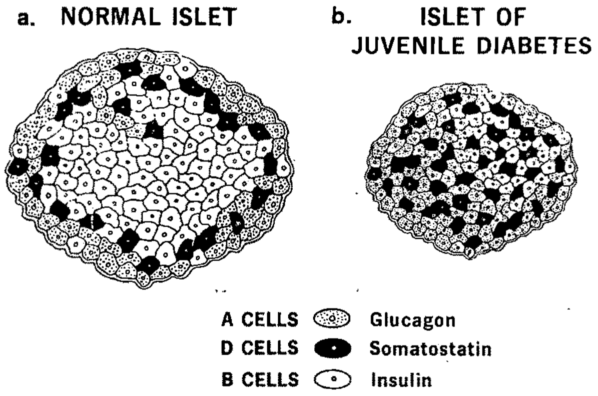


Figure 3 (*a*) The arrangement of the three major islet cell types in normal man. (*b*) The arrangement of islet cell types in juvenile-type diabetes.

somatostatin-like immunoreactivity of the islets (12–14). These cells make up approximately 10% of the total islet cell population. It should be noted that whereas the majority of β cells have contact only with other β cells, virtually all the α and D cells enjoy contact with at least one other cell type. This fact may be extremely important because each of the secretory products of these three cell types can influence the secretion of at least one of the two other cell types (Table 1). For example, somatostatin can inhibit the release of both insulin (15, 16) and glucagon (16, 17), glucagon can stimulate directly the release of insulin (18), insulin can inhibit directly the release of glucagon (19), and glucagon may possibly stimulate somatostatin release (20). In this tricellular area, the possibility for direct cell-to-cell action is maximal, since secretion products released into the intercellular channels would have instant contact with the surface of an adjacent cell. Tight junctions forming a network of ever-changing lines of fusion between the outer leaflets of adjacent cell membranes compartmentalize the intercellular space (Figure 4) and seem capable of enclosing secretory products (21). Local hormone-cell interactions could thereby be influenced by the creation of zones into which hormones can be preferentially sequestered, or from which they can be excluded. It should also be noted that the tricellular area receives the afferent nerves and blood vessels of the islets (22), and would consequently be the first region of the islets to receive signals incoming via these two routes. Finally, gap junctions, low-resistant pathways through which low–molecular weight substances may move from one cell to another cell without entering the extracellular space, and first reported to exist between like islet cells, i.e. β cells and β cells (21), have now been clearly established between α and β cells (23) and D and β cells; it is therefore highly likely that all of the various types of islet cells have intercellular communications with one another via gap junctions. In the functional sense, then, the entire islet, including the tricellular region, may be a syncytium. This may explain the precise coordination of glucagon and insulin secretion believed necessary for appropriate control of glucose homeostasis.

Table 1 Actions of islet-cell secretory products upon secretion by other islet cells

	β Cell	α Cell	D Cell
Insulin	—	↓	?
Glucagon	↑	—	↑
Somatostatin	↓	↓	—

ROLE OF THE GLUCAGON-SECRETING α CELLS IN DIABETES

The pathophysiologic model for diabetes mellitus, depicted in Figure 2B (21), ascribes the glucose underutilization of this disorder to insulin lack. (Although insulin lack also increases hepatic glucose production, it is believed that the in-

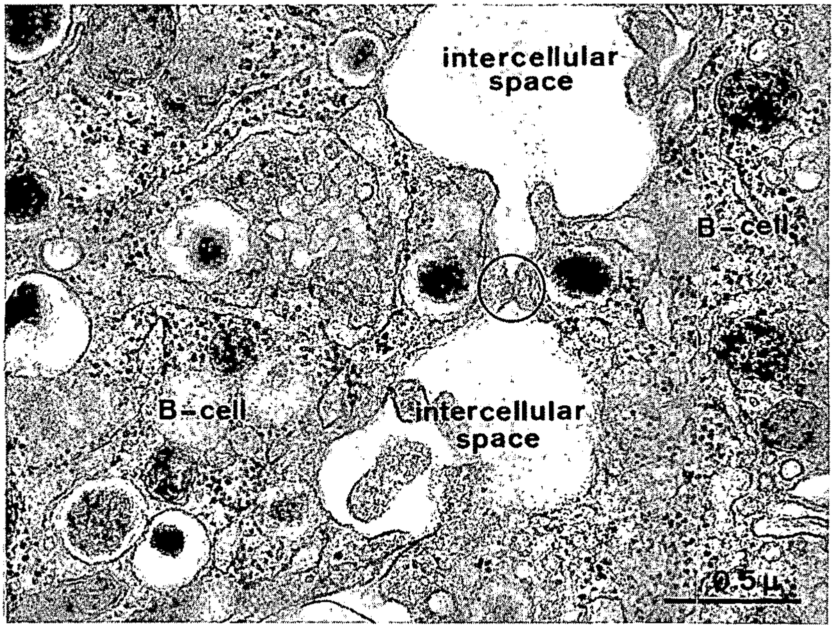


Figure 4 Thin-section electron photomicrograph showning points of focal contact between adjacent islet cells that seem to canalize the intercellular space.

creased glucose influx attributable to insulin lack per se is insufficient to cause the marked endogenous hyperglycemia that characterizes severe insulin deficiency and that, without glucagon, would result in endogenous hyperglycemia of only 150–170 mg/ml). Severe endogenous hyperglycemia, as encountered in uncontrolled diabetes and its marked hyperketonemia (3) requires both insulin lack and the presence of glucagon (or of substances closely resembling glucagon in their biologic actions on the liver). The essential role of glucagon in the pathogenesis of these two abnormalities (1) is supported by studies showing that somatostatin-induced blockade of pancreatic and/or extrapancreatic glucagon secretion can reduce hyperglycemia in insulin-deficient alloxan-diabetic dogs (24) or prevent its occurrence when insulin is discontinued in totally depancreatized dogs (25). Gerich and co-workers have demonstrated in juvenile diabetics that glucagon suppression by somatostatin blocks the severe endogenous hyperglycemia and the rise in β-hydroxybutyrate that otherwise occurs when insulin infusion is suddenly stopped and glucagon permitted to rise (2); replacement infusions of glucagon given together with the somatostatin restore the endogenous hyperglycemia and β-hydroxybutyrate levels to control values (26). In the diabetes of totally depancreatized humans, although aglucagonemia has been recorded by a radioimmunoassay employing immunoabsorbed plasma in its standard curves (27), conventional assay techniques reveal measurable immunoreactive glucagon (IRG) in the plasma of such patients (28), evidence that potentially bioactive IRG is present in the plasma of diabetics of all types. As shown in Table 2, whenever glucagon rises, as during an arginine infusion or a mixed meal, there is an increase in hyperglycemia, and whenever the rise in glucagon is blocked, either by arginine or insulin, no rise in glycemia occurs—in fact, when glucagon is reduced, glycemia declines (29). These findings, plus the fact that glucagon is increased relative to insulin levels in all known forms of endogenous hyperglycemia, ranging from stress hyperglycemia to experimental diabetes (5), provide support for the notion that glucagon is essential for the marked hepatic overproduction of glucose that characterizes the insulin-deficient state.

The abnormality in α-cell function that results in relative or absolute hyperglucagonemia clearly involves a loss of its glucose-sensing ability. Whereas hyperglycemia normally reduces plasma IRG and hypoglycemia increases it (30), in diabetics, a rise in glycemia, if unaccompanied by an increase in insulin, fails to lower plasma IRG (32) and will frequently cause a paradoxical increase (31), while hypoglycemia fails to elicit an increase in plasma IRG (32). Thus the main functional role of the α cell, that of preventing hypoglycemia, appears to have been relinquished in the juvenile diabetic patient and, ironically, inappropriately high levels of glucagon are present after meals—despite hyperglycemia, which in nondiabetics suppresses the glucagon response to protein ingestion (33).

A second identifiable functional abnormality of the α cell of the diabetic is hyperresponsiveness to aminogenic stimulation, manifested as an exaggerated rise in IRG during the infusion of arginine (34) or by a rise in response to a protein meal despite levels of hyperglycemia that in the nondiabetic would completely block a rise (33). It is not clear whether this response to aminogenic stimulation during hyperglycemia is another manifestation of loss of normal glucose-sensing ability or if it is a separate functional lesion. In juvenile diabetics (35, 36), but not in adult diabetics (36), high doses of insulin can reduce it to normal.

Around-the-clock observation of normal subjects reveals that postprandial IRG levels average 87 pg/ml, while in juvenile-type diabetics, even during optimal con-

Table 2 Relationship of changes in glucagon and glucose in insulinized human diabetics

	Δ Glucagon (pg/ml)	Δ Glucose mg/100 ml
Juvenile Diabetics		
Arginine infusion [Unger et al (34)]	↑ 325	↑ 80
Arginine infusion [Barnes & Bloom (27)]	↑ 300	↑ 59
Mixed meal plus insulin[a]	↑ 58	↑ 84
Mixed meal plus insulin plus somatostatin[a]	↓ 55	↓ 50
Totally Depancreatized Patients		
Arginine infusion [Barnes & Bloom (27)]	0	0
Arginine infusion [Palmer et al (28)]	↑ 61	↑ 48

[a] Gerich, J. E., Lorenzi, M., Karam, J. H., Schneider, V., Forsham, P. H. 1975. Abnormal pancreatic glucagon secretion and postprandial hyperglycemia in diabetes mellitus. *J. Am. Med. Assoc.* 234:159–65.

trol by aggressive insulin therapy, mean postprandial IRG levels average 118 pg/ml, significantly above the levels of normal patients—despite postprandial hyperglycemia averaging 155 mg/100ml, a level that in normal subjects would prevent a rise in plasma IRG (33).

Is the hectic IRG pattern, pictured in Figure 5, contributing to the irregular pattern of hyperglycemia, i.e. is this exogenous hyperglycemia, endogenous hyperglycemia, or both? The most likely answer is both. At least in that patient during stringent carbohydrate restriction to 30 g/day, endogenous hyperglycemia, perhaps mediated by the burst of glucagon secretion, contributed to the hyperglycemia and glycosuria. The 270-g/day reduction in carbohydrate intake failed to improve either his hyperglycemia or his glycosuria, suggesting that an increase in endogenously derived glucose influx approximately equal to the reduction in exogenously derived glucose influx took place (Figure 5).

However, when the plasma glucose of a diabetic can be normalized merely by reducing carbohydrate intake, one must assume that the contribution of endogenously derived glucose to the hyperglycemia is trivial compared to the contribution of dietary glucose. In mild adult-type diabetics who require no insulin and whose fasting plasma glucose levels are normal on dietary restriction alone, endogenous hyperglycemia can be assumed to be absent and therefore, in such patients glucagon is not contributing excessively to glucose influx; postprandial hyperglycemia is the result of reduced glucose utilization of ingested exogenous glucose secondary to inadequate insulin action. Indeed, there is accumulating evidence that in the milder hyperinsulinemic maturity-onset diabetic with hyperinsulinemia, glucagon levels, although high relative to the glucose concentration, are sensitive to further increases in hyperglycemia, i.e. that glucose sensing by the α cell is not completely abolished (37).

ETIOLOGY OF α-CELL MALFUNCTION

What is the etiology of the loss of the glucose-sensing capacity of the α cell in juvenile diabetes? Although it remains to be proven, there is reason to believe that the glucose sensor of the α cell may be the β cell. The evidence in favor of this follows below. Immunocytochemical studies of the islets of juvenile diabetics and streptozotocin-diabetic rat islets (8) reveal an absence of β cells, as schematized in Figure 3b. Instead, the islets consisted largely of glucagon-containing cells, somatostatin-containing D cells, and PP-containing cells. Instead of the narrow mantle of α cells of the nondiabetic islet, almost all of them in contact with β cells, in the juvenile diabetic islet the α cells of the juvenile diabetic are in contact largely with other α cells and have no contact with β cells. α cells of the canine gastric fundus and of glucagonomas (38, 39), which lack contact with β cells, also lack the capacity to respond to hyperglycemia, and in both instances their secretion is suppressed by an increase in insulin and stimulated by arginine. Severely diabetic α cells behave similarly: they, too, do not respond normally to glycemic change (33), are inhibited by insulin (40), and are stimulated by arginine (34). The insulin levels required to reduce the glucagon levels of juvenile diabetics to the normal range are above normal, a fact interpretable as evidence of α-cell insensitivity to insulin in diabetes

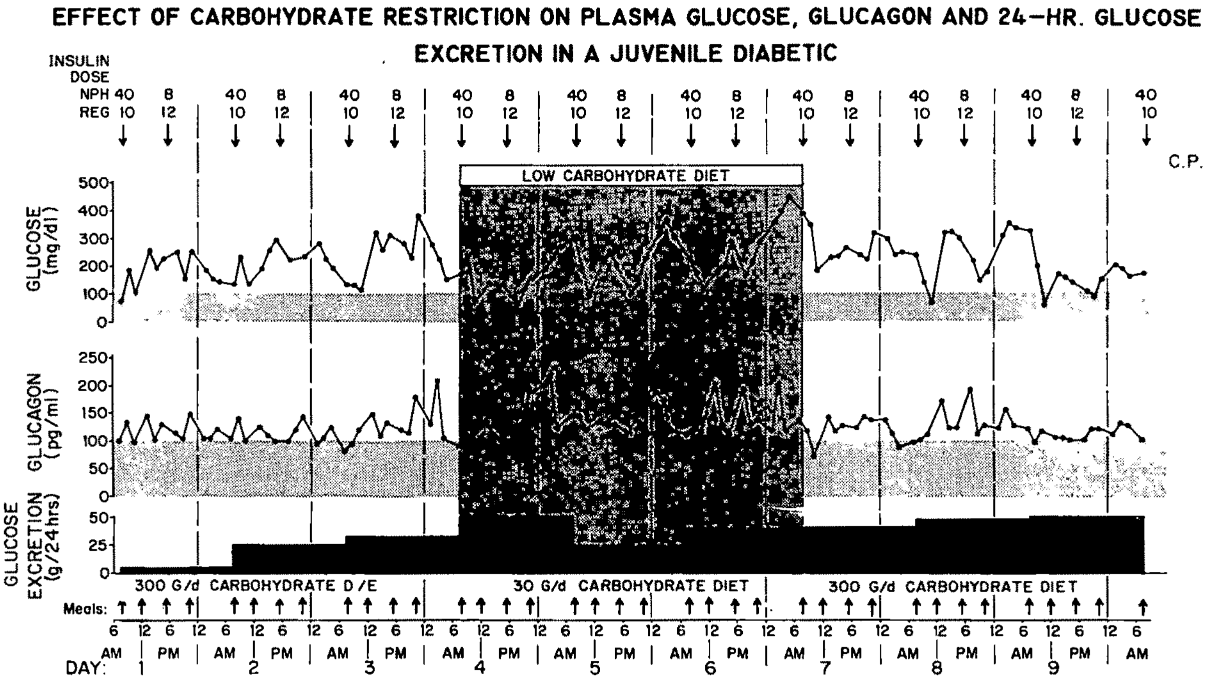


Figure 5 The pattern of plasma glucose and IRG levels in a typical juvenile-type diabetic while on a 300 g/day carbohydrate diet as well as a 30 g/day carbohydrate diet. Carbohydrate restriction failed to improve the hyperglycemia and glycosuria. Phasic rises in plasma IRG levels may have contributed to the hectic plasma-glucose fluctuations.

(40). However, equally valid is the view that insulin itself is the glucose-sensing signal, i.e. that insulin released in response to glycemic change from glucose-sensing β cells adjacent to the α cells makes instant contact with adjacent α cells and provides the signal for their suppression, as proposed by Weir (41). Rather than being insensitive to insulin, the α cells of the juvenile diabetic may in fact be exquisitely sensitive to insulin, responding to levels of insulin that, although pharmacologic in the systemic sense, fail to approach the insulin concentrations of the intercellular space separating α cells and β cells. Alternatively, other signals passing between β cells and α cells, perhaps via gap junctions, might provide the signal.

THERAPEUTIC IMPLICATIONS

Whatever the signal, there is evidence that a fundamental defect of the α cell, i.e. loss of glucose-sensing capacity, is present in the juvenile diabetic patient and cannot be corrected, at least not to normal, by normalizing plasma insulin levels. Although pharmacological levels of plasma insulin can suppress plasma IRG levels into or near normal range of normoglycemic nondiabetics, these insulin levels are abnormally high and must cause abnormally high glucose uptake by insulin-sensitive tissues. Such increased efflux of glucose from the extracellular space during insulin-induced suppression of glucagon, together with insulin's direct inhibiting effect on hepatic glucose production, must surely increase the risk of hypoglycemia. Therefore, in patients in whom endogenous glucose influx is clearly a contributing factor to the hyperglycemia, the use of a glucagon-suppressing agent that does not increase glucose utilization might be of value. A somatostatin-like agent might permit the administration of insulin in lower, more physiological doses designed only to maintain normal glucose efflux rather than to suppress glucagon; by independent control of hyperglucagonemia, normoglycemia might be achieved at lower insulin doses and lower and more physiological rates of glucose efflux, with less danger of hypoglycemia. This approach to the management of diabetic hyperglycemia will warrant careful scrutiny by those concerned with better management of diabetic hyperglycemia.

ACKNOWLEDGMENTS

This work was supported by NIH Grant AM 02700–16; Pfizer Laboratories, New York; Bristol Myers Company, New York; Mead Johnson Center, Evansville, Indiana; Dr. Karl Thomae GmbH, Germany; Hoffmann-LaRoche, Nutley, New Jersey; Hoechst Pharmaceutical Company, Somerville, New Jersey; CIBA-GEIGY Corporation, Ardsley, New York; the Upjohn Company, Kalamazoo, Michigan; Eli Lilly, Indianapolis, Indiana; 30K Rabbit Fund; grant 3.553.75 from Fonds National Suisse de la Recherche Scientifique; and by VA Institutional Research Support Grant 549-8000-01.

The authors wish to thank the following persons for outstanding technical work: Virginia Harris, Kay McCorkle, Margaret Bickham, Cathy Mitchell, Loretta Clendenen, Danny Sandlin, Sara Innis, and Helen Gibson.

For secretarial assistance, the authors thank Billie Godfrey and Susan Freeman.

Literature Cited

1. Unger, R. H., Orci, L. 1975. The essential role of glucagon in the pathogenesis of diabetes mellitus. *Lancet* 1:14–16
2. Gerich, J. E., Lorenzi, M., Bier, D. M., Schneider, V., Tsalikian, E., Karam, J. H., Forsham, P. H. 1975. Prevention of human diabetic ketoacidosis by somatostatin. Evidence for an essential role of glucagon. *N. Engl. J. Med.* 292:985–89
3. McGarry, J. D., Wright, P., Foster, D. 1975. Hormonal control of ketogenesis: Rapid activation of hepatic ketogenic capacity in fed rats by anti-insulin serum and glucagon. *J. Clin. Invest.* 55:1202–9
4. Cahill, G. F. Jr. 1971. Physiology of insulin in man. (Banting Mem. Lect.) *Diabetes* 20:785–99
5. Unger, R. H. 1976. Diabetes and the alpha cell. (Banting Mem. Lect.) *Diabetes* 25:136–51
6. Müller, W. A., Faloona, G. R., Unger, R. H. 1971. The influence of antecedent diet upon glucagon and insulin secretion. *N. Engl. J. Med.* 285:1450–54
7. Orci, L., Unger, R. H. 1975. Hypothesis: Functional subdivisions of the islets of Langerhans and the possible role of the insular D-cell. *Lancet* 2:1243–44
8. Orci, L., Baetens, D., Rufener, C., Amherdt, M., Ravazzola, M., Studer, P., Malaisse-Lagae, F., Unger, R. H. 1976. Hypertrophy and hyperplasia of somatostatin-containing D-cells in diabetes. *Proc. Natl. Acad. Sci.* 73:1338–42
9. Polak, J. M., Pearse, A. G. E., Grimelius, L., Bloom, S. R., Arimura, A. 1975. Growth hormone release inhibiting hormone in gastrointestinal and pancreatic D-cells. *Lancet* 1: 1220–22
10. Pelletier, G., Leclerc, R., Arimura, A., Schally, A. V. 1975. Immunohistochemical localization of somatostatin in the rat pancreas. *J. Histochem. Cytochem.* 23:699–701
11. Orci, L., Baetens, D., Dubois, M. P., Rufener, C. 1975. Evidence for the D-cell of the pancreas secreting somatostatin. *Horm. Metab. Res.* 7:400–2
12. Luft, R., Efendic, S., Hokfelt, T., Johansson, O., Arimura, A. 1974. Immunohistochemical evidence for the localization of somatostatin-like immunoreactivity in a cell population of the pancreatic islets. *Med. Biol.* 52: 428–30
13. Dubois, M. P. 1975. Immunoreactive somatostatin is present in discrete cells of the endocrine pancreas. *Proc. Natl. Acad. Sci.* 72:1340–43
14. Arimura, A., Sato, H., Dupont, A., Nishi, N., Schally, A. V. 1975. Somatostatin: Abundance of immunoreactive hormone in rat stomach and pancreas. *Science* 189:1007–9
15. Alberti, K. G. M. M., Christensen, N. J., Christensen, S. E., Hansen, A. P., Iversen, J., Lundbaek, K., Seyer-Hansen, K., Orskov, H. 1973. Inhibition of insulin secretion by somatostatin. *Lancet* 2:1299–1301
16. Mortimer, C. H., Turnbridge, W. M. G., Carr, D., Yeomans, L., Lind, T., Coy, D. H., Bloom, S. R., Kastin, A., Mallinson, C. N., Besser, G. M., Schally, A. V., Hall, R. 1974. Effects of growth hormone release-inhibiting hormone on circulating glucagon, insulin, and growth hormone in normal, diabetic, acromegalic and hypopituitary patients. *Lancet* 1:697–701
17. Koerker, D. J., Ruch, W., Chideckel, E., Palmer, J., Goodner, C. J., Ensinck, J., Gale, C. C. 1974. Somatostatin: Hypothalamic inhibitor of the endocrine pancreas. *Science* 184:482–83
18. Samols, E., Tyler, J. M., Marks, V. 1972. In *Glucagon. Molecular Physiology, Clinical and Therapeutic Implications*, ed. P. J. Lefebvre, R. H. Unger, p. 151. Amsterdam: Pergamon. 370 pp.
19. Samols, E., Harrison, J. 1976. Intraislet negative insulin-glucagon feedback. *Metabolism* 25:Suppl. 1, pp. 1443–47
20. Patton, G., Vale, W., Dobbs, R., Orci, L., Unger, R. H. 1976. Stimulation of pancreatic immunoreactive somatostatin (IRS) release by glucagon. *Metabolism* 25:Suppl. 1, p. 1449
21. Orci, L., Unger, R. H., Renold, A. E. 1973. Structural coupling between pancreatic islet cells. *Experientia* 29: 1015–18
22. Fujita, T. 1975. *Int. Symp. GEP Endocrine Syst. Kyoto, Jpn.*
23. Orci, L., Malaisse-Lagae, F., Ravazzola, M., Rouiller, C., Renold, A. E., Perrelet, A., Unger, R. H. 1975. A morphological basis for intercellular communication between A- and B-cells in the endocrine pancreas. *J. Clin. Invest.* 56:1066–70
24. Sakurai, H., Dobbs, R. E., Unger, R. H. 1974. Somatostatin-induced changes in insulin and glucagon secretion in nor-

mal and diabetic dogs. *J. Clin. Invest.* 54:1395–1402

25. Sakurai, H., Dobbs, R. E., Unger, R. H. 1975. The role of glucagon in the pathogenesis of the endogenous hyperglycemia of diabetes mellitus. *Metabolism* 24:1287–97
26. Gerich, J. E., Lorenzi, M., Bier, D. M., Tsalikian, E., Schneider, V., Karam, J. H., Forsham, P. H. 1976. Effects of physiologic levels of glucagon and growth hormone on human carbohydrate and lipid metabolism. Studies involving administration of exogenous hormone during suppression of endogenous hormone secretion with somatostatin. *J. Clin. Invest.* 47:875–84
27. Barnes, A. J., Bloom, S. R. 1976. Pancreatectomized man: A model for diabetes without glucagon. *Lancet* 1: 219–21
28. Palmer, J. P., Werner, P. L., Benson, J. W., Ensinck, J. W. 1976. Immunoreactive glucagon responses to arginine in three pancreatectomized humans. *Metabolism* 25:Suppl. 1, pp. 1483–85
29. Unger, R. H. 1976. Glucagon and blood sugar. *N. Engl. J. Med.* 294:1239
30. Unger, R. H. 1971. Glucagon physiology and pathophysiology: Abnormal alpha cell function as a characteristic of human diabetes mellitus. *N. Engl. J. Med.* 285:443–49
31. Buchanan, K. D., McCarroll, A. M. 1972. Abnormalities of glucagon metabolism in untreated diabetes mellitus. *Lancet* 2:1394–95
32. Gerich, J. E., Langlois, M., Noacco, C., Karam, J. H., Forsham, P. H. 1973. Lack of glucagon response to hypoglycemia in diabetes: Evidence for an intrinsic pancreatic alpha cell defect. *Science* 182:171–73
33. Müller, W. A., Faloona, G. R., Unger, R. H., Aguilar-Parada, E. 1970. Abnormal alpha cell function in diabetes. Response to carbohydrate and protein ingestion. *N. Engl. J. Med.* 283:109–15
34. Unger, R. H., Aguilar-Parada, E., Müller, W. A., Eisentraut, A. M. 1970. Studies of pancreatic alpha cell function in normal and diabetic subjects. *J. Clin. Invest.* 49:837–48
35. Gerich, J. E., Tsalikian, E., Lorenzi, M., Karam, J. H., Schneider, V., Gustafson, G., Bohannon, N. V. 1975. Normalization of fasting hyperglucagonemia and excessive glucagon responses to intravenous arginine in human diabetes mellitus by prolonged perfusion of insulin. *J. Clin. Endocrinol. Metab.* 41:1178–80
36. Raskin, P., Aydin, I., Unger, R. H. 1976. The effect of insulin on the exaggerated glucagon response to arginine stimulation in diabetes mellitus. *Diabetes* 25:227–29
37. Hatfield, H. H., Kalkhoff, R. K., Banasiak, M. F., Driscoll, T., Kim, H. J. 1974. Glucose suppression: Relation to pancreatic B-cell function. *Clin. Res.* 22:649A
38. Tiengo, A., Fedele, D., Marchiori, E., Nosadini, R., Muggeo, M. 1976. Suppression and stimulation mechanisms controlling glucagon secretion in a case of islet cell tumor producing glucagon, insulin, and gastrin. *Diabetes* 25:408–12
39. Boden, G., Owen, O. E., Quickel, K. E. 1976. Effect of chronic glucagon excess on glucose homeostasis and insulin secretion in man. *Diabetes* 25:Suppl. 1, p. 341
40. Raskin, P., Fujita, Y., Unger, R. H. 1975. Effect of insulin-glucose infusions on plasma glucagon levels in fasting diabetics and nondiabetics. *J. Clin. Invest.* 56:1132–38
41. Weir, G. C., Knowlton, S. D., Atkins, R. F., McKennan, K. X., Martin, D. B. 1976. Glucagon secretion from the perfused pancreas of streptozotocin-treated rats. *Diabetes* 25:275–82

Ann. Rev. Med. 1977. 28:131–41
Copyright © 1977 by Annual Reviews Inc. All rights reserved

FACTORS IN THE PATHOMECHANISM OF CHRONIC LYMPHOCYTIC LEUKEMIA

❖7230

Harald Theml, M.D., Richard Love, M.D.,
and Herbert Begemann, M.D.
Erste Medizinische Abteilung des Städtischen, Krankenhauses München-Schwabing, Munich, West Germany

INTRODUCTION

Clues as to etiology have not been forthcoming for chronic lymphocytic leukemia (CLL), as they have for the myeloid leukemias; therefore in CLL, studies have concentrated on the pathomechanism of the disease. There is a wealth of new data that characterize the pathognomonic cells in CLL and aspects of their production and turnover. Analysis of this information clarifies the pathomechanism of the clinical picture and offers a pathophysiological basis for the assessment of various therapeutic approaches.

CHARACTERISTICS OF THE PATHOGNOMONIC CELL IN CLL

Structural and Metabolic Features

For a long time it has been known that the lymphocytes in CLL are peculiar in being highly injury-prone, a phenomenon they manifest under the light microscope, where they appear as 'smudge' cells. Holt (1) attributed this sensitivity to an increased rigidity of the cell membrane. Under the electron microscope the preponderance of CLL cells strike one by their smooth nuclei and small mitochondria (2). Cytochemically, the glycogen content is markedly increased (3) and the glycogen turnover diminished (4). The enzymes β-glucuronidase and acid phosphatase are present in diminished amounts, a reflection of lower numbers of lysosomes (5). A specific marker chromosome, such as is found in chronic myeloid leukemia (CML), is generally lacking, although chromosomal abnormalities have been found in cases of CLL when multiple members of a family were affected (6). DNA metabolism in CLL lymphocytes is abnormal; diminished levels of desoxyribonuclease (7) and qualitatively altered DNA polymerases are found (7). RNA metabolism shows more significant deviations. Total RNA synthesis (8) and the RNA content of individual cells are increased (9), and nucleic-acid sequences not found in normal lymphocyte RNA are present (10). To date, evidence for RNA-instructed DNA polymerase (reverse transcriptase, revertase), like that in RNA tumor viruses, has been found in only a few cases of CLL (11).

Functional Abnormalities of CLL Lymphocytes

Our understanding of the dichotomy of the lymphatic system has come only in recent years. A brief presentation of this information follows, so that the subsequent details of alterations in CLL can be better appreciated.

NORMAL THYMUS-DERIVED AND BURSA-DERIVED DIVISION OF THE LYMPHATIC SYSTEM In the animal experiments of Sell and Gell it was first made clear that a portion of the lymphocytes carries immunoglobulin on its surface (12). Whereas in birds such lymphocytes are derived from the bursa fabricii (13) and were therefore decribed under the working name of B lymphocytes, in mammals they belong to a bone marrow–derived line (14). In healthy human beings, 22–36% of lymphocytes are found to have a cover of immunoglobulin on their cell surfaces by immunofluorescent staining or autoradiographic means (15). This suggests that every specific lymphoid population synthesizes and carries a heavy- and light-chain type that it secretes as antibody on stimulation with antigen and, in part, after morphological transformation to a plasma-cell population (14). B lymphocytes are further characterized by the presence of receptors for the Fc portion of aggregated IgG and by receptors for the activated third component of complement, which allow the binding of immunocomplement complexes (15).

From animal experiments it seems that the remaining two thirds of lymphocytes, which have no cell-surface immunoglobulin, are thymus-derived or at least acquire their mature functional capacity from the thymus (16), and are therefore designated T lymphocytes. With antigenic stimulation these lymphocytes turn into large blast cells (15), capable of giving off a number of factors called lymphokines (17). These lymphokines are important in the classical cell-mediated, delayed-hypersensitivity reaction. In addition, as "helper cells," T cells partially regulate B-cell activity (18). In vitro the transformation of T lymphocytes can be mediated through nonspecific mitogens. For example, three days after the addition of the plant protein phytohemagglutinin (PHA), 50–80% of T lymphocytes are transformed into blasts. A further specific property of T cells is their ability to form rosettes with sheep erythrocytes, in the absence of previous sensitization (15).

B-CELL PROPERTIES AND THEIR DISTURBANCES IN CLL The recent data showing an increased number of B cells in CLL seemed to render the clinical observation of a marked deficiency in all types of immunoglobulins somewhat incomprehensible. Wilson & Nossal (19) described patients in whom 87–93% of the lymphocytes demonstrated fluorescent staining for cell-surface immunoglobulins. In 14 of 20 cases, with over 80% immunoglobulin-positive lymphocytes, Grey, Rabellino & Pirofsky (20) were able to specify that the cell surface contained a heavy-chain class of type IgM, and either of two possible light chains. In these cases no monoclonal serum proteins were found, and the serum IgM levels were usually depressed. The percentage of immunoglobulin-positive cells was not correlated with disease activity or with the level of the lymphocytosis. In 25 patients, Aisenberg & Bloch (21) found the same preponderance of lymphocytes that apparently were carrying specific monoclonal IgM. Of particular interest was their observation that four patients had monoclonal IgM elevations in their serum. In one of these patients the light-chain type in the serum was identical to that on the lymphocyte cell surface. These observations have been confirmed by other authors (22, 23).

The ability to react with aggregated IgG as well as with immunocomplement complexes is interpreted as showing that the majority of CLL lymphocytes have binding sites for these B-cell markers. Thus three markers (cell-surface immunoglobulin, aggregated IgG, and immunocomplement complexes) allow 90% or more of CLL lymphocytes to be characterizable as B cells (19, 21, 24). There are, in addition, some unique features of these cells. Some cells are qualitatively different in having a decreased amount of detectable surface immunoglobulin (22, 23); some show a considerable diminution in the concentration of receptors for aggregated IgG (24). The phenomenon of "cap formation," a transient accumulation of immunoglobulin at one pole of the normal lymphocyte surface, is absent in CLL lymphocytes (22, 25). Finally, Cohnen et al (26) found crystalloid intracytoplasmic inclusions that might represent evidence of disrupted immunoglobulin production, most likely IgM aggregates.

All of these findings permit the interpretation that most of the lymphocytes in CLL have B-cell characteristics, with mostly IgM surface cover, disturbed membrane dynamics, and disturbed immunoglobulin secretion. Many authors consider these findings evidence of a neoplastic B-cell abnormality. Further support for this position is the evidence from stimulation tests with the plant pokeweed mitogen, which in vitro stimulates primarily B cells. Smith, Cowling & Barker (27) found a markedly reduced reaction of CLL lymphocytes to pokeweed mitogen. Others have demonstrated no reaction to pokeweed mitogen in an isolated CLL B-cell population, whereas a maximal reaction occurred with normal cells (28). Finally, however, Cohnen and his colleagues found, after pokeweed stimulation, a remaining fraction of 5–10% blast cells in which they could demonstrate electron-microscopically all the structures of normal immunoglobulin-producing B cells (29). It is therefore accepted that in CLL there is a small population of normal B cells with normal immunoglobulin-secreting ability mixed in with a majority of pathological B cells that cannot secrete immunoglobulins. This explains why total agammaglobulinemia hardly ever occurs, although the immunoglobulin levels are often so very low (30).

T-CELL FUNCTION IN CLL Among the major pathological B cells in CLL one finds T cells with likewise disturbed functional parameters. Studies with PHA and other T-cell–specific mitogens have shown that CLL lymphocytes frequently have a markedly diminished rate of transformation after stimulation in in vitro cultures (31–33). On the average, the values are 5–12% blasts on the third and fourth days with correspondingly low nucleoside incorporation. Most investigators have found a reciprocal relationship between the lymphocyte count in the peripheral blood and the rate of transformation (27, 31). Blasts induced by PHA from CLL lymphocytes are no different morphologically or electron microscopically from those induced from lymphocytes of normal persons (34). It is noteworthy that in cases of CLL with high lymphocyte counts, the relative maximal rate of blast transformation on PHA stimulation occurs somewhat later, on the fifth to seventh days (35). At a subcellular level, a lysosomal enzyme deficiency or the absence of an increased permeability of lysosomal membranes necessary for the extrusion of activated hydrolases, seems to be responsible for the poor response to PHA stimulation (36).

A quantitative determination of the number of T cells that form rosettes with sheep erythrocytes revealed a relatively low value of 2% in CLL, whereas normals show 25%. This decrease in rosette-forming T cells was found to be almost linearly and directly proportional to the extent of the lymphocytosis (37). Through gradient separation, Wybran, Chantler & Fudenberg (28) were able to collect a fraction of CLL lymphocytes, over 90% of which formed spontaneous rosettes. This fraction also carried no cell-surface immunoglobulin and showed normal response rates to PHA and pokeweed mitogen. This evidence suggests that there is a small, normal-reacting T-cell population in CLL. Clues as to the size of this population come from Aisenberg et al (38), who were able to mark 1–11% of CLL cells with antithymocyte serum. These values indicate a relative depression of T-cell numbers. In contrast, Catovsky and his colleagues calculated that in untreated, stable cases the absolute number of T cells was markedly elevated, while in rapidly advancing, often-treated cases, the number of T cells was depressed, even in absolute numbers (39).

It is necessary to qualify the foregoing data concerning B- and T-cell disturbances in CLL by noting that they are not valid for all cases. A very few cases (1–2%) have been described in which the ratios of B and T cells are exactly the reverse of these findings; here the disease has been called T-cell CLL (40). Because of its rarity we have dispensed with discussion of this special clinical entity here.

THE PROLIFERATION KINETICS OF LYMPHOCYTES IN CLL

The pathomechanism of the expansion of lymphoid cells into the blood, spleen, liver, and bone marrow seems to be the central question in CLL (41). Schiffer (42) suggested that an easily understood first parameter characterizing the mode of expansion is the accumulation rate of lymphocytes. In untreated patients, followed over years (42), this is found to be expressible as a straight line on a logarithmic graph, which describes an exponential increase of cell counts and a constant doubling time of cells in each phase of the disease. Even after therapeutic intervention, the increases in cell numbers that ensue are shown as lines of almost congruent steepness. The individual characteristic doubling times that can be calculated from this model are between 4 and 19 months in our patients and in those of Hansen (30). This is a relatively slow doubling time when compared with the values for lymph-node growth in non-Hodgkin's lymphomas (one month) (43) and Hodgkin's lymphoma (two months) (44). However, normal lymphatic proliferation results in no growth of the cell counts after attainment of adulthood through a balance of new-cell production and old-cell breakdown. Thus this relatively slow numerical increase of CLL cells represents a considerable disturbance. Obviously, this increase can come about through augmented production or diminished breakdown or both of these means. To clarify which of these alternatives may be operant, a number of isotopic studies of cell turnover have been carried out (45, 46). These have verified the slow turnover rate; thus, marking lymphocyte production with ^{3}H thymidine for long periods of time was necessary (47). The autoradiographic analyses offered the opportunity to separate short- and long-lived lymphocyte populations. From these studies we show in Table 1 the average values for CLL patients compared to those generally accepted to represent normal values from hematologically "healthy" cancer patients.

These mean values tend towards the higher end of the range of individual accumulation rates. A daily new production rate of 0.2–1%, or 300–1500 cells, is considered to be the usual range (48). The general problem of proliferation in CLL is clear from a study of this table:

1. In clinically apparent and treatment-requiring stages with elevated total cell counts, at least 15 times the normal number of lymphoid cells is produced daily.
2. Short-lived lymphocytes are increased 25-fold, and these cells account for most of the increase in total production. The average life span of these short-lived lymphocytes is in the normal range. This means that with increased production these cells accumulate at a moderately higher total level only.
3. An absolute increase in cell count is assured, however, by long-lived lymphocytes that have a lower rate of proliferation than the short-lived lymphocytes, but still have a rate ten times that of normal cells. Of particular importance is the five-fold increase in average life span of these long-lived cells.

This combination of increased production and delayed destruction leads to the exponential accumulation of these cells. In normal persons the daily production of five long-lived lymphocytes/μl, each having a life span of 365 days, is matched by a daily loss of the same quantity of cells, and a level of 5 times 365 $\approx$ 2000 cells/μl is maintained (48). In CLL patients, however, the daily production of 60

Table 1 Some characteristics of lymphocyte kinetics in normal patients and in patients with CLL (48, 47)

Factor	CLL	Normal
Lymphocytes (μl)	100,000	2,200
Production of lymphocytes (μl/day)	600	46
Short-lived lymphocytes (μl)	5,000	200
Production of short-lived lymphocytes (μl/day)	540	41
Average life span of short-lived lymphocytes (days)	5	6
Long-lived lymphocytes (μl)	95,000	2,000
Production of long-lived lymphocytes (μl/day)	60	5
Average life span of long-lived lymphocytes (years)	5	1

cells/μl, each having an average life span of five years, results in a level of 100,000 cells/μl.

The precursors of all these cells are primarily the blast cells of the lymphatic system; their initial products are large lymphocytes, and from these come small lymphocytes incapable of further division. It is clear, however, that not all blast cells proliferate, and there exists a relatively large resting/G_0 fraction (47). The production of lymphocytes in the bone marrow amounts to about one third that of the lymph nodes (47). It is interesting that the remaining white cells demonstrate a qualitatively normal maturation (49).

CIRCULATION KINETICS OF LYMPHOCYTES IN CLL

Since the studies of Gowans & Knight (50) and Schiffer (42), we have known that the division of lymphocytes and their easy exchange between the intra- and extravascular pools are decisive in their function. Upon leaving the intravascular space and permeating the tissues, especially those of the lymph nodes and spleen, the lymphocytes make contact with antigens and begin to carry out their specific functions. The path of recirculation leads from (*a*) the blood through the white pulp of the spleen and back into the blood, or (*b*) from the blood into the lymph-node cortices and from there by means of the efferent lymph circulation back into the blood, or (*c*) from the blood directly into the tissues, then by the afferent lymph channels into the regional lymph nodes and from these by the efferent lymph circulation back into the blood (50). Normally the extravascular readily accessible lymphocyte pool (RALP) is 25 times greater than the intravascular pool. Studies with ^{3}H cytidine-marked lymphocytes demonstrate considerable shift of lymphocytes toward the intravascular compartment in CLL: the RALP was found to be only 3–5 times greater than the intravascular circulating pool (51, 52). With each increase in the blood-lymphocyte count, the situation becomes more abnormal, with the accumulation of an even greater-than-normal fraction of cells in the intravascular circulation (51). Correspondingly, lymphocytes recirculating through the thoracic duct are relatively reduced in number, and are of proportionally different lymphocyte populations than those in the blood. Recirculating cells show normal T- and B-cell function (53). This means that the lymphocytes responsible for the clinical characteristics of CLL also have a pronouced disturbance in their ability to leave the vascular system and recirculate. Thus, on the one hand an intravascular accumulation of cells results, while on the other a complementary inability to migrate from the tissues leads to cell accumulation in organ compartments with resultant visceromegaly. Other studies (52) suggest that the surface properties of CLL lymphocytes, rather than anatomic alterations in vessels, are responsible for this inability to recirculate.

INTERACTION OF FACTORS IN THE PATHOMECHANISM OF CLL

To summarize, the following are the major disturbances of the lymphatic system in CLL:

1. Increased production of long-lived lymphocytes.
2. Hyperextended life span of long-lived lymphcytes.
3. Increased production of short-lived lymphocytes with normal elimination.
4. A disturbance of exchange of lymphocytes between intra- and extravascular pools and vice versa.
5. Loss of function of the majority of B-type lymphocytes.

A causal connection of factors responsible for these disturbances is difficult as long as kinetic features of specific lymphoid cells cannot be correlated directly with immunological-function studies. However, on quantitative grounds there can be no doubt that the pathologically functionless B cells, which dominate the usual clinical pictures as they constitute over 90% of cells, are long-lived lymphocytes. By contrast, the population of lymphocytes that turns over rapidly can be relatively heterogeneous. On the one hand, the rate of proliferation of this population corresponds to that of a population of plasma cells from which it can be separated on morphological grounds (47), so that the balance of cells can be assumed to be normal B cells. On the other hand, this heterogeneous population could contain the remaining T-cell population, which is also increased in volume (39); has normal ability to recirculate; and whose cells Rieke & Schwarz (54) found, in animal experiments, to have a relatively short life span.

The central question is why a normal equilibrium between cell proliferation and breakdown is not maintained in the dominant cell population of long-lived and functionless elements, and why instead these cells have an increased rate of production and an increased life span. Unfortunately, the mediators of feedback between cell breakdown and cell production are not adequately defined for even normal lymphopoiesis (55). Various pieces of evidence (56) suggest that normal regulation is dependent on (*a*) normal cell-membrane properties, (*b*) normal cell products (such as immunoglobulins of B cells and lymphokines of T cells) and (*c*) normal exchange of intra- and extravascular pools. Profound disturbances occur in each of these in CLL.

If one must hypothesize a causal chain in CLL, it most likely begins, as Galton suggests (57), with the malignant degeneration of a cell clone for unknown reasons. This results in functionless cells with abnormal surface-membrane properties and pathological metabolism. The lack of function is primarily responsible for the extremely long life of the cells, which together characterize CLL as "an accumulative disease of immunologically incompetent lymphocytes" (58). Beyond this CLL is, as demonstrated, also a disease of proliferation. This could be based on the fact that the pathological cells give off no normal catabolites, either as nucleotides (59), or as functional cell products, as self-regulating inhibitors of proliferation (60). It is true, however, that after massive destruction of cells following radiotherapy or after cytotoxic therapy, no compensatory increase in the doubling time is observed (61). These results speak against a direct regulation of proliferation by cell-adherent mediators or by secreted mediators related to the cell count. Another possibility is that the inability to secrete immunoglobulins causes a loss of a braking mechanism on proliferation normally applied after reaching a certain level of antibody produced by specific lymphocytes (62). However, this seems unlikely because doubling times remain constant through normal and almost agammaglobulinemic phases in the course of the illness (30).

The disturbance of recirculation would seem to be of secondary importance. However, Vincent & Gunz (63) believe that the detector responsible for maintaining normal lymphocyte counts is in the postcapillary venule (which must be passed in the course of recirculation) and that the recognition site is the PHA-receptor of lymphocytes. Absence of this recognition site and less frequent passage by the detector would lead to a loss of normal regulation. On the other hand, other data show (51) that even when there is a considerable decrease in absolute numbers, more or at least as many cells recirculate as in hematologically healthy people (64). Thus an adequate number of signals must come to the detector.

In all, no model-explanations of the dysregulation of the lymphatic system in CLL remain unchallenged. Thus for the present we are left with the suggestion of neoplastic and autonomous cells, which through loss of function, like most neoplastic cells, have a long life span and a relatively low rate of proliferation. The only peculiarity

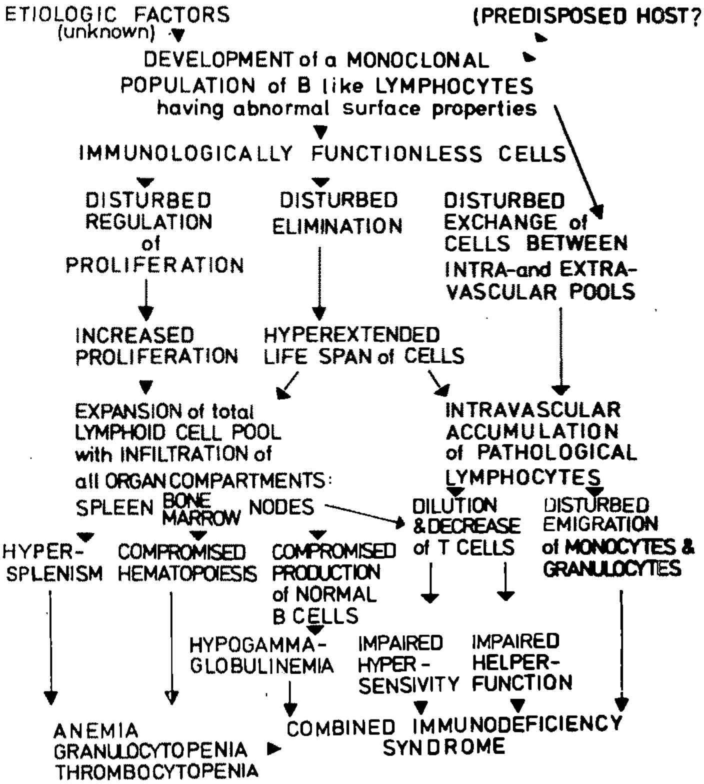

ETIOLOGIC FACTORS (unknown)
(PREDISPOSED HOST?)
DEVELOPMENT of a MONOCLONAL POPULATION of B like LYMPHOCYTES having abnormal surface properties
IMMUNOLOGICALLY FUNCTIONLESS CELLS
DISTURBED REGULATION of PROLIFERATION
DISTURBED ELIMINATION
DISTURBED EXCHANGE of CELLS BETWEEN INTRA-and EXTRA-VASCULAR POOLS
INCREASED PROLIFERATION
HYPEREXTENDED LIFE SPAN of CELLS
EXPANSION of total LYMPHOID CELL POOL with INFILTRATION of all ORGAN COMPARTMENTS:
SPLEEN BONE MARROW NODES
INTRAVASCULAR ACCUMULATION of PATHOLOGICAL LYMPHOCYTES
DILUTION & DECREASE of T CELLS
DISTURBED EMIGRATION of MONOCYTES & GRANULOCYTES
HYPER-SPLENISM
COMPROMISED HEMATOPOIESIS
COMPROMISED PRODUCTION of NORMAL B CELLS
HYPOGAMMA-GLOBULINEMIA
IMPAIRED HYPER-SENSIVITY
IMPAIRED HELPER-FUNCTION
ANEMIA GRANULOCYTOPENIA THROMBOCYTOPENIA
COMBINED IMMUNODEFICIENCY SYNDROME

Figure 1 Pathways in the pathomechanism of CLL.

in CLL, then, is the apparently high-degree of differentiation of the degenerate end cell (65). An outline of the pathomechanism in CLL, with the principal factors presented here, can be seen in Figure 1.

CONCLUSION

It has been shown that CLL is characterized by a piling up of highly differentiated lymphocytic cells. These cells have the structural and metabolic characteristics of a neoplastic cell line of B lymphocytes (except in cases of "T-cell CLL"). However, they lack the immunoglobulin-secreting ability of normal B cells, and are immunologically incompetent and inert. Next to this population, there is a normal but reduced population of B cells and a periodically slightly increased T-cell population. The accumulation of pathological cells is based on a 10-fold increase in proliferation of cells that have a 5-fold increase in their life span. In addition, there is a disturbance of exchange of cells between the intra- and extravascular pools. These characteristics clarify the development of the clinical picture: through packing of the bone marrow with pathological cells on the one hand, and the spleen on the other, anemia, thrombocytopenia, and finally granulocytopenia develop. The gradual displacement of normal B cells often leads to extreme hypogammaglobulinemia as the main component of a multifactorial syndrome of immune deficiency.

Literature Cited

1. Holt, P. J. L., Pal, S. G., Catovsky, D., Lewis, S. M. 1972. Surface structure of normal and leukaemic lymphocytes. I. Effect of mitogens. *Clin. Exp. Immunol.* 10:555–70
2. Schrek, R. 1972. Ultrastructure of blood lymphocytes from chronic lymphocytic and lymphosarcoma cell leukemia. *J. Natl. Cancer Inst.* 48:51–64
3. Quaglino, D., Cowling, D. C., Hayhoe, F. G. J. 1964. Cytochemical and autoradiographic studies on normal leukaemic cells in shortterm tissue cultures. *Br. J. Haematol.* 10:417–36
4. Stjernholm, R. L., Dimitrov, N. V., Zito, S. 1970. Carbohydrate metabolism in leukocytes. XIII. Differentiation by metabolism of leukemic leukocytes into three groups. *J. Reticuloendothel. Soc.* 7:539–50
5. Douglas, S. D. 1972. Electron microscopic and functional aspects of human lymphocyte response to mitogens. *Transplant. Rev.* 11:39–59
6. Gunz, F. W., Fitzgerald, P. H., Adams, A. 1962. An abnormal chromosome in chronic lymphocytic leukaemia. *Br. Med. J.* 2:1097–99
7. Loeb, L. A., Poiesz, B. J., Williams, R. O., Springgate, C. 1973. On multiple DNA-polymerase after stimulation of normal and leukemic lymphocytes: In *Proc. Leukcocyte Culture Con., 7th,* p. 151. New York: Academic
8. Mukherjee, A. B., Waite, R. G., Cohen, M. M., Bernstein, R. 1972. Incorporation of uridine–^{3}H and sodium acetate – ^{14}C in lymphocytes derived from normal and leukemic individuals. *Cancer Res.* 32: 1833–36
9. Laszlo, J., Huang, A. T. F., Kremer, W. B. 1970. Studies on human leukemic cells and normal lymphocytes. In *Methods in Cancer Research,* ed. H. Busch, K. Smetana, 5:373
10. Neimann, P. E., Henry, P. H. 1969. RNA-DNA-hybridization and hybridization-competition studies of rapidly labeled RNA from normal and chronic lymphocytic leukemia lymphocytes. *Biochemistry* 8:275–82
11. Spiegelman, S. 1974. Molecular evidence for viral agents in human cancer and its chemotherapeutic consequences. *Cancer Chemother. Rep.* 58:595–613
12. Sell, S., Gell, P. G. H. 1965. Studies on rabbit lymphocytes in vitro. I. Stimulation of blast transformation with an antiallotype serum. *J. Exp. Med.* 122: 423–40
13. Rabellino, E., Grey, H. M. 1971. Immunoglobulins on the surface of lym-

phocytes. III. Bursal origin of surface immunoglobulins on chicken lymphocytes. *J. Immunol.* 106:1418–20

14. Cooper, M.D., Lawton, A. R., Kincade, P. W. 1972. A two-stage model for development of antibody-producing cells. *Clin. Exp. Immunol.* 11:143–49
15. Bentwich, Z., Kunkel, H. G. 1973. Specific properties of human B and T lymphocytes and alterations in disease. *Transplant. Rev.* 16:29–50
16. Schlesinger, M. 1970. Anti-theta antibodies for detecting thymus-dependent lymphocytes in the immune response of mice to SRBC. *Nature* 226:1254–56
17. Wolstencroft, R. A., Mathew, M., Oates, C., Maini, R. M., Dumonde, D.C. 1971. Lymphocyte mitogenic factor in cell-mediated immunity. In *The Role of Lymphocytes and Macrophages in the Immunological Response,* ed. D.C. Dumonde. Berlin-Heidelberg-New York: Springer. 100 pp.
18. Greaves, M., Janossy, G. 1972. Elicitation of selective T and B lymphocyte responses by cell surface binding ligands. *Transplant. Rev.* 11:87–130
19. Wilson, J. D., Nossal, G. J. V. 1971. Identification of human T and B lymphocytes in normal peripheral blood and in chronic lymphocytic leukaemia. *Lancet* 2:788–91
20. Grey, H. M., Rabellino, E., Pirofsky, B. 1971. Immunoglobulins on the surface of lymphocytes. IV. Distribution in hypogammaglobulinemia, cellular immune deficiency and chronic lymphatic leukemia. *J. Clin. Invest.* 50:2368–75
21. Aisenberg, A. C., Bloch, K. J. 1972. Immunoglobulins on the surface of neoplastic lymphocytes. *N. Engl. J. Med.* 287:272–76
22. Huber, Ch., Kurz, R., Asamer, H., Huber, H., Braunsteiner, H. 1974. Die Differenzierung menschlicher Blutlymphocyten mit serologischen und autoradiographischen Methoden. *Klin. Wochenschr.* 52:127–33
23. Ternynck, T., Dighiero, G., Follezou, J., Binet, I. L., 1974. Comparison of normal and CLL lymphocyte surface Ig determinants. Using peroxidase-labeled antibodies. I. Detection and quantitation of light chain determinants. *Blood* 43:789-95
24. Preud'homme, J. L., Seligmann, M. 1972. Surface bound immunoglobulins as a cell marker in human lymphoproliferative diseases. *Blood* 40:777–94
25. McLaughlin, H., Wetherly-Mein, G., Pitcher, C., Hobbs, J. R. 1973. Non-immunoglobulin-bearing "B" lymphocytes in chronic lymphatic leukaemia. *Br. J. Haematol.* 25:7–14
26. Cohnen, G., Douglas, S. D., König, E., Bremer, K., Brittinger, G. 1973. Crystalloid intracytoplasmic inclusions in lymphocytes of a patient with chronic lymphocytic leukemia. *Mt. Sinai J. Med. NY* 40:249–59
27. Smith, J. L., Cowling, D. C., Barker, C. R. 1972. Response of lymphocytes in chronic lymphocytic leukaemia to plant mitogens. *Lancet* 1:229–33
28. Wybran, I., Chantler, S., Fudenberg, H. H. 1973. Isolation of normal T cells in chronic lymphocytic leukaemia. *Lancet* 1:126–29
29. Cohnen, G., Douglas, D., König, E., Brittinger, G. 1973. Pokeweed mitogen response of lymphocytes in chronic lymphocytic leukemia: a fine structural study. *Blood* 42:591–600
30. Hansen, M. M. 1973. Chronic lymphocytic leukemia. *Scand. J. Haematol. Suppl.* 18:1–282
31. Elves, M. W., Collinge, M., Israels, M. C. 1967. The potential of lymphocytes from patients with leukaemia and reticuloses to transform under the influence of phytohaemagglutinin. *Acta Haematol.* 37:100–8
32. Hirschhorn, K., Schreibman, R. R., Bach, F. H., Sitzbach, L. E. 1964. Invitro studies of lymphocytes from patients with sarcoidosis and lymphoproliferative diseases. *Lancet* 2:842–43
33. Schrek, R., Rabinowitz, Y. 1963. Effects of phytohemagglutinin on rat and normal and leukemic human blood cells. *Proc. Soc. Exp. Biol. Med.* 133:191–94
34. Clausen, K. P., Bouroncle, B. A. 1969. The ultrastructure of phytohaemagglutinin (PHA) stimulated lymphocytes of chronic lymphatic leukaemia. *Blood* 34:179–90
35. Havemann, K., Rubin, A. D. 1968. The delayed response of chronic lymphocytic leukemia lymphocytes to phytohaemagglutinin in vitro. *Proc. Soc. Exp. Biol. Med.* 127:668–71
36. König, E., Brittinger, G., Cohnen, G. 1973. Relation of lysosomal fragility in CLL lymphocytes to PHA reactivity. *Nature New Biol.* 244:247–48
37. Fröland, S. S. 1972. Binding of sheep erythrocytes to human lymphocytes. A probable marker of T lymphocytes. *Scand. J. Immunol.* 1:269–80
38. Aisenberg, A. C., Bloch, K. J., Long, J. C., Colvin, R. B. 1973. Reaction of nor-

mal human lymphocytes and chronic lymphocytic leukemia cells with an antithymocyte antiserum. *Blood* 41: 417–23

39. Catovsky, D., Miliani, E., Okos, A., Galton, D. A. G. 1974. Clinical significance of T-cells in chronic lymphocytic leukaemia. *Lancet* 2:751–52
40. Brout, J. C., Flandrin, G., Sasportes, M., Preud'homme, J. L., Seligmann, M. 1975. Chronic lymphocytic leukaemia of T-cell origin. *Lancet* 2:890–93
41. Begemann, H., Rastetter, J., Kaboth, W. 1975. *Klinische Hämatologie.* Stuttgart: Thieme. 887 pp.
42. Schiffer, L. M. 1968. Kinetics of chronic lymphocytic leukemia. *Ser. Haematol.* 1:3–23
43. Iversen, O. H., Iversen, U., Ziegler, J. L., Blumin, A. Z. 1974. Cell kinetics in Burkitt lymphoma. *Eur. J. Cancer,* 10:155–63
44. MacDonald, J. S., Laugier, A., Schlienger, M. 1970. In *Progress in Lymphology II,* ed. M. Viamonte, pp. 16–19. Stuttgart: Thieme
45. Hamilton, L. D. 1956. Nucleic acid turnover studies in human leukaemic cells and function of lymphocytes. *Nature* 178:597–99
46. Zimmermann, T. S., Godwin, H. A., Perry, S. 1968. Studies of leukocyte kinetics in chronic lymphocytic leukemia. *Blood* 31:277–91
47. Theml, H., Trepel, F., Schick, P., Kaboth, W., Begemann, H. 1973. Kinetics of lymphocytes in chronic lymphocytic leukemia: Studies using continuous ^{3}H-thymidine infusion in two patients. *Blood* 42:623–36
48. Schick, P. 1975. Lymphozytenkinetik bei lymphatischen Systemerkrankungen. In: *Lymphozyt und Klinische Immunologie,* ed. H. Theml, H. Begemann, pp. 90–95, Berlin-Heidelberg-New York: Springer. 220 pp.
49. Theml, H., Schick, P., Trepel, F., Heltzel, U., Kaboth, W., Fliedner, T. M., Begemann, H. 1974. Kinetik der nicht-lymphatischen Leukozyten bei chronischer lymphatischer Leukämie. *Blut* 28:217 (Abstr.)
50. Gowans, J. L., Knight, E. J. 1964. The route of re-circulation of lymphocytes in the rat. *Proc. R. Soc. London.* 159:257–82
51. Bremer, K., Fliedner, T. M., Schick, P. 1973. Kinetic differences of autotransfused ^{3}H-cytidine labeled blood lymphocytes in leukemic and non-leukemic lymphoma patients. *Eur. J. Cancer* 9:113–24
52. Manaster, J., Frühling, J., Stryckmans, P. 1973. Kinetics of lymphocytes in chronic lymphocytic leukemia. I. Equilibrium between blood and a "readily accessible pool." *Blood* 41:425–38
53. Flad, H. D., Huber, C., Bremer, K., Menne, H. D., Huber, H. 1973. Impaired recirculation of B lymphocytes in chronic lymphocytic leukemia. *Eur. J. Immunol.* 3:688–93
54. Rieke, W. O., Schwarz, M. R. 1967. The types of rat thoracic duct lymphocytes which respond to phytohemagglutinin in vitro. *Acta Haematol.* 38:121–28
55. Wolstenholme, G. E. W., Knight, J. 1973. Immunopotentiation. *Ciba Found. Symp. No. 18.* Amsterdam: North-Holland, Elsevier
56. Brent, L., Davies, A. J. S. 1975. Die Regulation des lymphatischen Systems. See Ref. 48, pp. 41–47
57. Galton, D. A. G. 1966. The pathogenesis of chronic lymphocytic leukemia. *Can. Med. Assoc. J.* 94:1005
58. Dameshek, W. 1967. Chronic lymphocytic leukemia—an accumulative disease of immunologically incompetent lymphocytes. *Blood* 29:566–84 (Suppl.)
59. Cooper, H. L. 1973. Lymphocyte RNA labelling with ^{3}H-uridine. See Ref. 7, p. 119
60. Weiss, P., Kavanau, J. L. 1958. A model of growth and growth control in mathematical terms. *J. Gen. Physiol.* 41:1–48
61. Trepel, F., Schick, P. 1976. Proliferation und Wachstum von lymphatischen Zellpopulationen bei malignen Lymphomen. *Hämatol. Blüttransfus.* 18:In press
62. Uhr, J. W., Finkelstein, M. S. 1967. The kinetics of antibody formation. *Prog. Allergy* 10:37–83
63. Vincent, P. C., Gunz, F. W. 1970. Control of lymphocyte level in the blood. *Lancet* 2:342–44
64. Brass, B., Theml, H., Backmund, H., Lob, G., Seifert, J., Brendel, W., Spelsberg, F. 1973. Drainage des Ductus thoracicus beim Menschen. *Med. Klin.* 68:1399–404
65. Davis, S., 1975. Hypothesis: Differentiation of the human lymphoid system based on cellsurface markers. *Blood* 45:871–80

Ann. Rev. Med. 1977. 28:143–52
Copyright © 1977 by Annual Reviews Inc. All rights reserved

PAGET'S DISEASE OF BONE

❖7231

Will G. Ryan, M.D.
Department of Medicine, Rush Medical College, Section of Endocrinology and Metabolism, Rush Presbyterian-St. Luke's Medical Center, Chicago, Illinois 60612

Paget's disease of bone was described almost exactly a century ago (1876) by an English surgeon, Sir James Paget, after whom the disorder is named. He presented to the Royal Medical and Chirugical Society a case description of a man whom he had followed for twenty years until his death and autopsy and who was affected by "a rare form of chronic inflammation of bones." He proposed to call it after its most striking character "osteitis deformans." Since that time it has been the subject of numerous papers: it has been described in a skull from the Gallo-Roman era, a parietal bone from an Egyptian tomb, a femur from neolithic man, and several paleopathological specimens of American Indians. No convincing examples of the disorder have been found in animals other than humans. According to Barry, there are striking racial and geographical differences in its distribution, it being relatively more common in the United Kingdom, central Europe, and Australia and New Zealand; relatively less common in North America; somewhat less so in Scandinavia and southern Europe; and relatively rare in Africa, the Middle East, India, Japan, and China. The work of Schmorl (1932) in Germany and Collins (1956) in England on autopsy material indicate that in both countries about 3–4% of the population over age 40 have some evidence of the disease (1.6% in persons of all ages) (1, 2). Roentgenologic studies in living subjects confirm these findings and suggest an incidence of 1% in the fifth decade, which rises steadily to approximately 10% in the ninth. Sex ratios appear to be about equal, with perhaps a slight preponderance of males (1, 2).

ETIOLOGY

Many theories of etiology have been advanced since Paget described the disease as a chronic inflammatory process, although most of them have only a historical interest: The origin has been thought to be infectious, syphilitic, toxic, or neurotropic. Endocrine glands have been blamed: the "absence of perversion of some internal secretion," or more specifically, the pituitary, the parathyroids, or an imbalance between the parathyroids and adrenals. It has been thought of as a disease of mineral metabolism or as a local disturbance in bone metabolism. None has withstood the test of time and no one offers the explanation for the basic process, i.e. "the anarchic destruction of bone without respect for the existing architecture accompanied by an extraordinary increase in all the elements of the local vasculature forming the periosteal plexus" (1). Several authors have suggested that Paget's is a primary vascular disease. The inflammatory theory was revived in the 1950s when Albright & Henneman found suppression of features of the disease by cortisone (3) and later by aspirin (4).

The role of heredity has been stressed. Cases have been reported in identical twins, and in four generations in one instance. As early as 1883, a case had already been reported of a woman with characteristic clinical features whose father had been known to have similar bowing of the limbs, and since then an impressive number of articles have appeared in the literature calling attention to affected families (1). McKusick, after careful review of material reported, suggested likely transmission of the disorder by a simple autosomal Mendelian-dominant gene (5). We have suggested that the disorder might be neoplastic (6) on the basis of its histological characteristics and response to an inhibitor of RNA synthesis (mithramycin). Recently, Rebel et al have suggested that the disorder may be induced by a virus. Electron microscopic examination of bone from patients with Paget's disease has revealed that the nuclei of the osteoclasts contain an inclusion body that resembles the papovavirus (7).

The presence of angioid streaks in some patients with Paget's disease, as well as the occasional association with pseudoxanthoma elasticum, has suggested that it may be a more generalized disorder of connective tissue (5). Francis & Smith found that the amount and stability of a polymeric collagen fraction extracted from the skin of patients with Paget's disease is significantly reduced compared with matched controls (8). Misra has reported that the hydroxylysinorleucine cross-link in bone collagen from patients with Paget's disease was considerably increased (9). This may, however, be due to increased formation of embryonic or woven bone due to the rapid turnover of bone in Paget's disease, since this cross-link is more characteristically found in less mature bone. Franck et al found that the incidence of rheumatic problems distinct from those attributable to Pagetic involvement was "surprising" (10).

CLINICAL FEATURES

Excellent reviews on large series have covered the clinical features of Paget's disease and, except for a brief description and recent developments, these are not described here. The concept of excessive osteoclastic destruction of bone followed by compensatory osteoblastic repair, resulting in greatly increased bone turnover in affected areas in Paget's disease, propounded by Albright in the 1940s, is still valid today (11). Increased osteoclastic destruction of the collagen matrix of bone results in an elevated urinary hydroxyproline with the highest values being found in this disorder (20–30 times normal is commonly found in urine of patients with extensive active disease). Krane et al reported in 1967 that patients with Paget's disease excrete polypeptides in the urine that have amino-acid composition and other properties resembling those of fragments of collagen. The pattern of isotope incorporation in vivo suggested that the fragments are derived from collagen that has been synthesized and rapidly degraded, or that they were rapidly synthesized but not incorporated into tropocollagen molecules. This nondialyzable material, however, accounted for only about 10% of the total hydroxyproline (12).

Genuth & Klein studied a unique patient with simultaneous idiopathic hypoparathyroidism and active Paget's disease, and demonstrated that endogenous parathyroid hormone (PTH) is not essential for bone resorption in this disease. Treatment with exogenous PTH resulted in restoration of serum calcium and phosphorus to normal and enhanced bone resorption, as evidenced by increased urinary hydroxyproline and calcium excretion. A similar response to exogenous PTH was observed in a euparathyroid patient with Paget's disease in whom severe transient hypercalcemia was also produced (13). These findings are in accord with previous observations that parathyroidectomy had little effect in a patient with extensive active Paget's disease (11).

Increased vascularity is characteristic of Pagetic bone, and results in increased blood flow and warmth of the skin overlying bone near the surface. This had been attributed to arteriovenous fistulae, but Rhodes et al, using radioisotope-labeled microspheres, found no vessels of a size to suggest this and concluded that the increased vascularity was of normal type (14).

Heistad et al studied regulation of blood flow by comparing blood flow in "Pagetic" extremities to flow in the contralateral normal ones and found a mean increase in blood flow approximately seven times that of the normal. The contribution of cutaneous flow to the increase was evaluated with epinephrine iontophoresis, which suppresses flow to skin but not to underlying tissue. Epinephrine iontophoresis decreased blood flow in Pagetic extremities to near normal, suggesting that the increase in blood flow in Pagetic extremities is primarily the result of cutaneous vasodilatation (15).

Recent developments in bone scanning using either ^{18}F- or ^{99}Tc-labeled ethane hydroxyl diphosphonate (EHDP) have facilitated evaluation of the anatomic distribution of the Pagetic lesions. Shirazi et al described some of the characteristic scan abnormalities as follows: in long bones the lesion begins at one or both ends, spreading down the shaft only when the ends are involved full-width with progression stopping at the limits of bone. Involvement on both sides of a joint was coincidental, and isolated diaphysial involvement was not seen. There was a propensity to preserve rather than obliterate normal bony outlines, as in fibrous dysplasia metastatic disease, osteomyelitis, etc. Only 10% presented with monostotic involvement, with the majority being in skull, pelvis, and spine, in contrast to Barry's finding monostotic lesions primarily in long bones. Visualization of the entire skeleton by scan has resulted in only a slight difference in anatomic distribution analysis provided by others. Pelvic lesions were seen in 78%, spine in 63%, skull in 48%, femur in 48%, scapula in 37%, tibia in 22%, and humerus in 17%, with lesser amounts in other bones. In the spine, lumbar vertebrae were most frequently involved, and 60% had both spine and pelvis involvement. In only 2% was the axial skeleton totally spared. In long bones, proximal involvement was greater than distal. Facial bone involvement was uncommon, with mandibular involvement being relatively more common (16). Khairi et al found that sclerotic lesions on radiographs were always asymptomatic, and concluded that the ^{18}F bone scan detects clinically active lesions more often (17).

TREATMENT

The past decade, for Paget's disease, has been characterized by revolutionary developments in its treatment. Three modalities, calcitonins, cytotoxic agents, and diphosphonates, have been used, with generally favorable results for all with regard to lowering serum alkaline phosphatase and urinary hydroxyproline, and for symptomatic relief. One of the most striking effects seen has been the regression of neurologic abnormalities (paraplegia, etc) that have presumably been due to relief of neural compression with treatment. Regression of lesions on bone scan has been observed, as well as reversion of radiographic lesions toward normal.

Calcitonins

Natural porcine, and synthetic salmon and human calcitonins have all been used to treat Paget's disease; the studies reported within the last few years are too numerous to cite all of them (18–25). Porcine calcitonin is the least potent of the three in the human, and its use has largely been abandoned. Galante et al measured the acute calcium-lowering effect of the three in patients with Paget's disease and found that salmon calcitonin produced approximately the same effect as a ten-times-greater dose of human calcitonin. Porcine calcitonin, however, was least potent and, compared with human calcitonin, required an approximately tenfold dose for a similar response (26).

A common observation with calcitonin treatment has been a reduction in signs of activity with a plateau after 6 to 9 months and a rebound toward pretreatment values after 1–2 years. Mechanisms proposed for this phenomenon have been the development of neutralizing antibodies, secondary hyperparathyroidism due to lowering of serum calcium by calcitonin (27), and a direct escape of the bone cells from the inhibitory effects of calcitonin (28). Dube et al (1973) found very low titers of circulating anticalcitonin antibodies after 2 years of treatment (29). Haddad & Caldwell, however, found that the development of neutralizing antibodies coincided with rebound clinical and biochemical indices of relapse (30). Woodhouse et al used synthetic human calcitonin to help avoid the development of neutralizing antibodies (31). Rojanasathit, Rosenberg & Haddad later reported a biochemical response to human calcitonin in patients who had become resistant to salmon calcitonin (32). Burchhardt et al found no evidence of secondary hyperparathyroidism in patients treated with salmon calcitonin (33). According to Raisz et al, the most important cause of the rebound effect may be the escape phenomenon, which is exhibited by osteoclasts under continuous calcitonin treatment in vitro (28). This concept was affirmed by Wright et al, who found a loss of specific calcitonin receptors in bone under the influence of calcitonin treatment (34).

Side effects of calcitonin have been minor skin rashes, nausea, facial flushing, and swollen hands, but in none of the studies were patients withdrawn from calcitonin because of adverse reactions.

Bone biopsies have indicated that calcitonin therapy results in normal new-bone formation (31), and radiological evidence of regression has been seen with long-term treatment with human calcitonin, which was particularly evident in a 6-year-old boy with juvenile Paget's disease (35–37).

Grimaldi et al reported the lack of effect on deafness secondary to Paget's disease in 10 patients treated for a mean of 3 months with salmon calcitonin (38). Moffatt reported improvement in hearing in one patient treated with calcitonin. However, the lack of radiological evidence of skull disease in the patient leaves the relation of the deafness to Paget's disease in doubt (39).

Avramides et al reported beneficial effects of treatment with weekly injections of salmon calcitonin (40), and later reported that symptomatic and biochemical evidence of improvement after cessation of treatment with salmon calcitonin was maintained in 10 of 13 patients for up to one year (41).

From the information available, calcitonin seems to be a very safe drug for long-term use, but its effectiveness over the several decades of disease duration expected in Paget's disease remains to be seen.

Diphosphonates

Diphosphonates are compounds structurally related to pyrophosphate, but resistant to degradation by pyrophosphatase. They bind strongly to crystals of hydroxyapatite, block the growth and dissolution of such crystals in vitro, and slow bone turnover rate in experimental animals. Reports of the clinical use of diphosphonates have been confined so far to sodium editronate (EHDP, disodium ethane-1-hydroxy-1,1-diphosphonate). (42–45). Results that are both quantitatively and temporally similar to those obtained with calcitonins, on both biochemical parameters and symptoms, have been seen with effectiveness that is dose-related (oral administration 1–20 mg $(kg)^{-1}$ $(day)^{-1}$. During administration there is a reversible increase in plasma phosphate at higher doses, which is unexplained. Orally administered and free of apparent symptomatic side effects, they appeared to be ideal for treatment of Paget's disease. However, Smith et al (45, 46) have reported biopsy evidence of accumulation of osteoid after several months of therapy, suggesting impairment of mineralization of both normal and diseased bone. Kantowitz et al reported that a high proportion of patients receiving the higher dose (20 mg/kg) became markedly worse with severe pain over lesions and 7 of 29 patients who received 10–20 mg/kg/day developed 10 fractures, mostly at sites of increased radiolucency in long bones. They concluded that although EHDP suppresses Pagetic bone resorption, poorly mineralized repair bone may be structurally unsound, contribute to pain, and be subject to fracture (47). Currently, evaluation of treatment with lower doses is being conducted, but the reported ill effects leave the usefulness of long-term administration of at least sodium editronate in doubt.

Cytotoxic Drugs

Mithramycin and actinomycin D are antibiotics that bind to DNA, and in so doing block RNA synthesis, presumably exerting their cytotoxic effect in this manner. During the course of our search for a cytotoxic drug that might be effective against the (presumed neoplastic) excessive osteoclastic activity seen in Paget's disease, we encountered the report of Parsons et al (48), describing the hypocalcemia observed in patients being treated with mithramycin and assumed that the observed effect might be due to osteoclast toxicity. Prior to that, we had treated two Pagetic patients with actinomycin D in low dosage (0.1–0.2 mg/day X 5 days) and had observed transient effects on alkaline phosphatase. [Actinomycin D was used because of its reported blocking of the effect of vitamin D and parathyroid hormone on bone (49, 50)]. We observed dramatic effects of the administration of mithramycin (25 μg $(kg)^{-1}(day)^{-1}$ X 10 days) on serum alkaline phosphatase and urinary hydroxyproline accompanied by symptomatic relief of pain and regression of signs of neurological compression (6, 51, 52). Our further experience in approximately 150 patients and as reported by others (53–55) has been confirmatory of these observations, and objective evidence of remission remains for one to two years in most, and up to seven years in one. The prime disadvantage of mithramycin has been its hepato-, nephro-, and platelet toxicity, and fatalities directly attributable to the drug have been observed in patients in whom it was being used for the treatment of cancer. (56). However, we have seen no published reports of serious toxicity in patients being treated for Paget's disease in whom generally lower doses (15–25 μg/kg) are used, in contrast to those in cancer chemotherapy. (25–50 μg/kg). In our experience hepatotoxicity, as reflected in increases in serum enzymes, has been mild and transient (usually reversible within 48 hr) and without residue. Nephrotoxicity [blood urea nitrogen (BUN) 40–60 mg/100 ml] has been slightly more prolonged (approx. 7–10 days), but infrequent ($<$ 10%) and three patients have had BUN's remaining in a slightly elevated range (20–28 mg%). The only observed significant thrombocytopenia (75,000/cm) was when a patient received 50 μg/kg for 5 days through error. This, too, was transient and unaccompanied by other significant adverse effects (on the patient, not the investigator). Interestingly, this patient did not have a more prolonged remission of activity of her Paget's disease. We have never observed leukopenia secondary to mithramycin, but did on the only occasion that we have administered actinomycin D in full therapeutic dose (0.5 mg/day X 5 days), which did not have as significant or prolonged effect as mithramycin. Fenelly et al (57, 58) have made the most extensive comparisons between mithramycin and actinomycin D and find their effects to be generally similar with, however, a less profound hypocalcemic effect of actinomycin D. Subsequent studies have generally been directed at achieving therapeutic effects with lessened toxicity, generally by using smaller doses or less frequent administration (59–61). Elias et al reported beneficial effects from the administration of 25 μg of mithramycin every 2–3 weeks for several months (55). Our own experience indicates that objective effects are reasonably closely related to dose used and frequency of administration (62). We have maintained a few patients on weekly injections of mithramycin for up to three years without evidence of cumulative toxicity. Our personal experience with a variety of other cytotoxic agents has not been fruitful.

MISCELLANEOUS TREATMENTS

Condon (63) and Christiansen (64) reported beneficial effects of glucagon infusions and Sekel reported persistent relief of pain with calcium infusion, although biochemical improvement was only temporary (65). Phosphate (66) and fluoride (67, 68) have been used in past years, but have been largely abandoned with the advent of more definitely effective agents. Hosking et al reported enhanced effects of combined sodium editronate and synthetic human calcitonin (69), and our own experience indicates some degree of enhancement of the effect of mithramycin.

CONCLUSIONS

Paget's disease affects approximately 3% of the population over age forty. It may be genetically transmitted (autosomal-dominant), but recent evidence suggests a viral etiology. Effective agents for control of the disorder have been developed over the past decade, but all have significant drawbacks. Calcitonin tends to lose its effectiveness in significant numbers of patients over a few months, mithramycin has significant toxicity, and diphosphonates tend to cause osteomalacia.

ACKNOWLEDGMENTS

J. W. Bastian, PhD, of Armour Pharmaceutical Company provided many of the references. Ms. Karen Lee Williams provided expert assistance in manuscript preparation.

Literature Cited

1. Deuxchaines, C. N., Krane, S. M. 1964. Paget's disease of bone: Clinical and metabolic observations. *Medicine Baltimore* 43:233–66
2. Barry, H. C. 1969. *Paget's Disease of Bone.* Edinburgh & London: Livingston. 196 pp.
3. Albright, F., Henneman, P. H. 1955. The suppression of Paget's disease with ACTH and cortisone. *Trans. Assoc. Am. Physicians* 68:238–46
4. Maurice, P. F., Lynch, T. N., Bastomsky, C. H., Dull, T. A., Avioli, L. V., Henneman, P. H. 1962. Metabolic evidence for suppression of Paget's disease of bone by aspirin. *Trans. Assoc. Am. Physicians* 75:208–16
5. McKusick, V. A. 1972. *Heritable Disorders of Connective Tissue,* p. 718. St. Louis: Mosby. 4th ed.
6. Ryan, W. G., Schwartz, T. B., Perlia, C. P. 1969. Effects of mithramycin on Paget's disease of bone. *Ann. Intern. Med.* 70:549–57
7. Rebel, A., Malkani, K., Basle, M. 1974. Particularités ultra structurales des osteoclastes de la maladie de Paget. *Rev. Rhum. Mal. Osteo-Articulaires* 41:767–71
8. Francis, M. J. O., Smith, R. 1974. Evidence of a generalized connective tissue defect in Paget's disease of bone. *Lancet* 1:841–42
9. Misra, D. P. 1975. Crosslink in bone collagen in Paget's disease. *J. Clin. Pathol.* 28:305–8
10. Franck, W. A., Bress, N. M., Singer, F. R., Krane, S. M. 1974. Rheumatic manifestations of Paget's disease of bone. *Am. J. Med.* 56:592–603
11. Albright, F., Reifenstein, E. 1948. *The Parathyroid Glands and Metabolic Bone Disease.* Baltimore: Williams & Wilkins
12. Krane, S. M., Munoz, A. J., Harris, E. D. 1967. Collagen like fragments: Ex-

cretion in urine of patients with Paget's disease of bone. *Science* 157:713–15

13. Genuth, S. M., Klein, L. 1972. Hypoparathyroidism and Paget's disease: the effect of parathyroid hormone administration. *J. Clin. Endocrinol. Metab.* 35:693–99
14. Rhodes, B. A., Greyson, N. D., Hamilton, C. R. Jr., White, R. I., Giargiana, F. A., Wagner, H. M. 1972. Absence of anatomic arteriovenous shunts in Paget's disease of bone. *N. Engl. J. Med.* 287:444–50
15. Heistad, D. D., Abboud, F. M., Schmid, P. G., Mark, A. L., Wilson, W. R. 1975. Regulation of blood flow in Paget's disease of bone. *J. Clin. Invest.* 55:69–74
16. Shirazi, P. H., Ryan, W. G., Fordham, E. W. 1974. Bone scanning in evaluation of Paget's disease of bone. *Crit. Rev. Clin. Radiol. Nucl. Med.* 5:523–58
17. Khairi, M. R. A., Wellman, H. N., Robb, J. A., Johnston, C. C. 1973. Paget's disease of bone (osteitis deformans): Symptomatic lesions and bone scan. *Ann. Intern. Med.* 79:348–51
18. DeRose, J., Singer, F. R., Avramides, A., Flores, A., Dziadiw, R., Baker, R. K., Wallach, S. 1974. Response of Paget's disease to porcine and salmon calcitonins: Effects of long term treatment. *Am. J. Med.* 56:858–66
19. Greenberg, P. B., Dogle, F. H., Fisher, M. T., Hillyard, C. J., Joplin, G. F., Pennock, J., MacIntyre, I. 1974. Treatment of Paget's disease of bone with synthetic human calcitonin. *Am. J. Med.* 56:867–70
20. Kanis, J. A., Horn, D. B., Scott, R. D. M., Strong, J. A. 1974. Treatment of Paget's disease of bone with synthetic salmon calcitonin. *Br. Med. J.* 3:727–31
21. Hamilton, C. R. 1974. Effects of synthetic salmon calcitonin in patients with Paget's disease of bone. *Am. J. Med.* 56:315–22
22. Avramides, A., Baker, R., Wallach, S. 1974. Metabolic effects of synthetic salmon calcitonin in Paget's disease of bone. *Metabolism* 23:1037–46
23. Wallach, S., Avramides, A., Flores, A., Bellavia, J., Cohn, S. 1975. Skeletal turnover and total body elemental composition during extended calcitonin treatment of Paget's disease. *Metabolism* 24:745–53
24. Shai, F., Baker, R. K., Wallach, S. 1971. The clinical and metabolic effects of porcine calcitonin on Paget's disease of bone. *J. Clin. Invest.* 50:1927–40
25. Haddad, J. G., Birge, S. J., Avioli, L. 1970. Effects of prolonged thyrocalcitonin administration on Paget's disease of bone. *N. Engl. J. Med.* 283:549–55
26. Galante, L., Joplin, G. F., MacIntyre, I., Woodhouse, N. J. Y. 1973. The calcium lowering effect of synthetic human porcine and salmon calcitonin in patients with Paget's disease. *Clin. Sci.* 44:605
27. Dube, W. J., Goldsmith, R. J., Arnaud, S. B., Arnaud, C. B. 1971. Hyperparathyroidism secondary to long term therapy of Paget's disease of bone with calcitonin. *Clin. Res.* 19:371
28. Mundy, G. R., Raisz, L. G. 1974. Drugs for disorders of bone: Pharmacological and clinical considerations. *Drugs* 8:250–89
29. Dube, W. J., Goldsmith, R. S., Arnaud, S. B., Arnaud, C. D. 1973. Development of antibodies to porcine calcitonin during treatment of Paget's disease of bone. *Mayo Clin. Proc.* 48:43–46
30. Haddad, J., Caldwell, J. G. 1972. Calcitonin resistance: Clinical and immunologic studies in subjects with Paget's disease of bone treated with porcine and salmon calcitonins. *J. Clin. Invest.* 51:3133–41
31. Woodhouse, N. J. Y., Reiner, M., Bordier, P., Kalu, D. N., Fisher, M., Foster, G. V., Joplin, G. F., MacIntyre, I. 1971. Human calcitonin in the treatment of Paget's bone disease. *Lancet* 1:1139–43
32. Rojanasathit, S., Rosenberg, E., Haddad, J. 1974. Paget's bone disease: Response to human calcitonin in patients resistant to salmon calcitonin. *Lancet* 2:1412–15
33. Burckhardt, P. M., Singer, F. R., Potts, J. J. 1973. Parathyroid function in patients with Paget's disease treated with salmon calcitonin. *Clin. Endocrinol.* 2:15–22
34. Wright, D. R., Ivey, J. L., Tasjian, A. H. Jr. 1976. Self induced loss of calcitonin receptors in bone: A possible explanation for "escape". *Clin. Res.* 24:461A
35. Woodhouse, N. J. Y., Joplin, G. F., MacIntyre, I., Doyle, F. H. 1972. Radiological regression in Paget's disease treated by human calcitonin. *Lancet* 2:992–94
36. Doyle, F. H., Pennock, J., Greenberg, P. B., Joplin, G. F., MacIntyre, I. 1974. Radiological evidence of a dose-related response to long term treatment of Pa-

get's disease with human calcitonin. *Br. J. Radiol.* 47:4–8

37. Doyle, F. H., Woodhouse, N. J. Y., Glen, A. C. A., Joplin, G. F., MacIntyre, I. 1974. Healing of the bones in juvenile Paget's disease treated by human calcitonin. *Br. J. Radiol.* 47:9–15
38. Grimaldi, P. M. G. B., Mohamedally, S. M., Woodhouse, N. J. Y. 1975. Deafness in Paget's disease: Effect of salmon calcitonin treatment. *Br. Med. J.* 1:726
39. Moffatt, W. H., Morrow, J. D., Simpson, N. 1974. Effects of calcitonin therapy in deafness associated with Paget's disease of bone. *Br. Med. J.* 4:203
40. Avramides, A., Flores, A., DeRose, J., Wallach, S. 1975. Treatment of Paget's disease of bone with once a week injections of salmon calcitonin. *Br. Med. J.* 2:632
41. Avramides, A., Flores, A., DeRose, J., Wallach, S. 1976. Paget's disease of the bone: Observations after cessation of long term synthetic salmon calcitonin treatment. *J. Clin. Endocrinol. Metab.* 42:459–63
42. Smith, R., Russell, R. G. G., Bishop, M. 1971. Diphosphonates and Paget's disease of bone. *Lancet* 1:945–47
43. Smith, R., Russell, R. G. G. 1972. Treatment of Paget's disease with a diphosphonate (sodium editronate). *Sem. Drug Treat.* 2:77–81
44. Altman, R. D., Johnston, C. C., Khairi, M. R. A., Wellman, H., Serafini, A. N., Sankey, R. R. 1973. Influence of disodium editronate on clinical and laboratory manifestations of Paget's disease of bone (osteitis deformans). *N. Engl. J. Med.* 289:1379–84
45. Russell, R. G. G., Smith, R., Preston, C., Walton, R. J., Woods, C. G. 1974. Diphosphonates in Paget's disease. *Lancet* 1:894–98
46. Smith, R., Russell, R. G. G., Bishop, M. 1973. Paget's disease of bone: Experience with a diphosphonate (disodium editronate) in treatment. *Q. J. Med.* 166:235–56
47. Kantrowitz, F. G., Byrne, M. H., Krane, S. M. 1975. Clinical and metabolic effects of the diphosphonate EHDP in Paget's disease of bone. *Clin. Res.* 23:445A
48. Parsons, V., Baum, M., Self, M. 1967. Effects of mithramycin on calcium and hydroxyproline metabolism in patients with malignant disease. *Br. Med. J.* 1:474–77
49. Eisenstein, R., Passavoy, M. 1964. Actinomycin D inhibits parathyroid hormone and Vitamin D activity. *Proc. Soc. Exp. Biol.* 117:77
50. Rasmussen, H., Arnaud, C., Hawker, C. 1964. Actinomycin D and the response to parathyroid hormone. *Science* 154:1019
51. Ryan, W. G., Schwartz, T. B., Northrop, G. 1970. Experiences in the treatment of Paget's disease of bone with mithramycin. *J. Am. Med. Assoc.* 213:1153–57
52. Ryan, W. G., Schwartz, T. B., Northrop, G. 1972. Treatment of Paget's disease with mithramycin: Further experiences. *Sem. Drug Treat.* 2:57–64
53. Condon, J. R., Reith, S. B. M., Nassim, J. R., Millard, F. J. C., Hilb, A. 1971. Treatment of Paget's disease of bone with mithramycin. *Br. Med. J.* 1:421–23
54. Russell, A. S., Lentle, B. C. 1974. Mithramycin therapy in Paget's disease. *Can. Med. Assoc. J.* 110:397–400
55. Elias, E. G., Evans, J. T. 1972. Mithramycin in the treatment of Paget's disease of bone. *J. Bone Jt. Surg.* 54:1730–36
56. Kennedy, B. J. 1970. Metabolic and toxic effects of mithramycin during tumor therapy. *Am. J. Med.* 48:494–503
57. Fennelly, J. J., Groarke, J. F. 1971. Effect of actinomycin D on Paget's disease of bone. *Br. Med. J.* 1:423–26
58. Fennelly, J. J. 1971. Clinical and biochemical studies of Paget's disease of bone with emphasis on the effects of RNA inhibitors actinomycin D and mithramycin. *Ir. J. Med. Sci.* 140:431–48
59. Aitken, J. M., Lindsay, R. 1973. Mithramycin in Paget's disease. *Lancet* 1:1177–78
60. Somerville, P. J., Evans, R. A. 1975. Actinomycin D in the treatment of Paget's disease of bone. *Med. J. Aust.* 2:13–16
61. Ajlouni, K., Theil, G. B. 1975. Mithramycin effects on calcium, phosphorous and parathyroid hormone in osseous Paget's disease. *Am. J. Med. Sci.* 269:13–18
62. Lebbin, D., Ryan, W. G. 1974. Outpatient treatment of Paget's disease of bone with mithramycin. *Ann. Intern. Med.* 81:635
63. Condon, J. R. 1971. Glucagon in the treatment of Paget's disease of bone. *Br. Med. J.* 4:719–21
64. Christiansen, C., Tonnesen, K. H. 1974. Zinc-protamine glucagon in the treatment of Paget's disease of bone. *Acta Med. Scand.* 196:495–96

65. Sekel, R. 1973. Calcium infusion in Paget's disease. *Lancet* 1:372–73
66. Goldsmith, R. S. 1972. Treatment of Paget's disease with phosphate. *Sem. Drug Treat.* 2:69–73
67. Riggs, B. L., Jowsey, J. 1972. Treatment of Paget's disease with fluoride. *Sem. Drug Treat.* 2:65–75
68. Avioli, L. V., Berman, M. 1968. Role of magnesium metabolism and the effects of fluoride therapy in Paget's disease of bone. *J. Clin. Endocrinol. Metab.* 28:700–10
69. Hosking, D. J., Bijvoet, D. L., Van Aken, J., Will, E. J. 1976. Paget's bone disease treated with diphosphonate and calcitonin. *Lancet* 1:615–16

Ann. Rev. Med. 1977. 28:153–57
Copyright © 1977 by Annual Reviews Inc. All rights reserved

MALIGNANT HYPERPYREXIA

❖7232

C. Ronald Stephen, M.D., C.M., F.F.A.R.C.S.
Department of Anesthesiology, Washington University School of Medicine,
St. Louis, Missouri 63110

About ten years ago, the syndrome now known as malignant hyperthermia became recognized as a distinct entity in anesthetic practice (1–4). Because of the associated high mortality and the worldwide occurrence of this syndrome (over 600 cases are now known), interest in its etiology, prevention, and treatment has continued apace. In 1971 an International Symposium on Malignant Hyperthermia was held in Toronto, Canada (5).

Early recognition of this syndrome has been and is a stumbling block to its proper therapy. The typical patient is a child or young adult, previously healthy, in whom a rapidly developing increase in body temperature occurs over a short span of time (10–30 min). If body temperature is not being monitored, the first signs to be noted may be a hot skin or the development of a tachycardia and/or arrhythmias. By this time the body temperature may be found to be 39–42°C. At this point, therapy must be considered just as much of an emergency as that for a cardiac arrest, if the life of the patient is to be saved. The situation is similar to that of a patient suffering from heat stroke.

The etiological factors responsible for the development of malignant hyperthermia are still unknown. It is known, however, that this syndrome is based on a hereditary predisposition that is transferred as an autosomal dominant trait (6–8). It is also firmly established from clinical experience that this syndrome can be triggered in susceptible individuals by the muscle relaxant succinylcholine and/or a potent inhalation anesthetic (e.g. cyclopropane, ethyl ether, halothane, methoxyflurane, enflurane). In a majority of patients (about 75%), there is an associated hypertonus or muscle rigidity that develops with onset of the syndrome, and it is the consensus of opinion at present that malignant hyperthermia develops because of a lesion occurring in skeletal muscle.

The reasons for implicating the muscle at all in the pathogenesis are not conclusive, but reasonably suggestive. Caffeine is known to cause contracture in isolated muscle fibers by enhancing the release and inhibiting the uptake of calcium ions by the sarcoplasmic reticulum. Kalow and colleagues (9) have shown that muscle preparations from survivors of malignant hyperpyrexia were more sensitive to caffeine rigor than normal muscle cells, and that the same result was found with exposure to the triggering agent halothane. It can be postulated that, at least in rigid forms of malignant hyperthermia, the triggering agent releases calcium from a defective calcium-storing sarcoplasmic reticulum membrane of the muscle cell. The resulting high myoplasmic calcium in turn activates phosphorylase kinase, thus accelerating glycolysis; activates myosin ATPase, thus accelerating hydrolysis of ATP to ADP, phosphate, and heat; inhibits troponin, permitting contraction to occur; and is taken up into the mitochondria, causing an uncoupling of oxidative phosphorylation. These events are capable of liberating sufficient heat to account for the fulminant rise in body temperature (10, 11).

Animal models (Landrace, or Poland China pigs) have been described in which malignant hyperpyrexia can be triggered (12, 13). Although the stress syndrome in pigs may not be identical to that seen in humans (14), experiments with these animals have confirmed many observations in humans and have initiated specific types of therapy for application to humans. It was in this manner that procaine was suggested as a method of successfully treating the established syndrome (15). The rationale for such therapy is that the caffeine-induced rigor in muscle can be blocked by procaine, due to a stabilizing effect on the sarcoplasmic reticulum membrane, and it was suggested that a similar result might be obtained with triggering agents such as halothane. More recently it has been shown that dantrolene sodium, a muscle relaxant that interferes with calcium release from the sarcoplasmic reticulum, can in vitro and in vivo slow or reverse the malignant nature of this syndrome in susceptible swine (16–18).

PRESENT STATUS AND THERAPY

The incidence of malignant hyperpyrexia today is estimated to be from 1:50,000 to 1:100,000. In the early years of its recognition the mortality rate was 70–80%, the patients dying in cardiac failure from "heat stroke" or from the sequelae of the runaway metabolism associated with the severe pyrexia. Now the mortality rate is believed to be 30–40%, the decrease being attributed to perhaps better detection of susceptible patients, earlier recognition of the syndrome as it is developing, and more alert and definitive treatment once the syndrome is diagnosed.

Since this disorder is genetically related, patients seen preoperatively are queried carefully about anesthetic problems they or their relatives may have had previously. As more susceptible families are being identified, the members of the families are being warned and medi-alert bracelets are being worn.

For several years it was believed that a preanesthetic elevated creatine phosphokinase (CPK) level denoted a susceptible patient. However, it is now realized that this test is relatively nonspecific (19, 20), and that a high serum CPK reading may only be of prognostic value when it is found in an individual who is a member of a family known to be susceptible to the syndrome (21). It would appear that a careful history is the best way of detecting susceptible patients in advance of anesthesia administration.

Once administration of anesthesia is begun, an abnormal reaction to the injection of succinylcholine intravenously should alert one to a possibly susceptible patient. If muscle fasciculations are greater than anticipated, and the patient fails to relax as expected, serious consideration should be given to postponing the operation until further investigation can be conducted. Prior administration of a small dose of D-tubocurarine to modify the fasciculations does not necessarily protect a susceptible patient.

Probably the best way to detect the onset of the syndrome, which may begin shortly after the beginning of anesthesia or which may be delayed for one or more hours, is by monitoring body temperature from the outset. Monitoring may be via the rectum, esophagus, nasopharynx, or tympanic membrane (22). A constantly increasing body temperature for no apparent reason should be of significant concern. In a short time tachypnea, tachycardia, or possible arrhythmias will develop, and the blood at the operative site will become darker than normal. The patient also will often show evidence of hypertonus and rigidity.

When such a situation develops, an emergency must be declared and vigorous therapy begun at once. The operation is delayed or concluded within a matter of minutes. Total efforts are made to cool the patient by surrounding him in ice, administering cold intravenous fluids (Ringer's lactate) at a rapid rate, instituting ice-water enemas and gastric lavage, and washing out any open body cavities with cold saline. Because metabolism is proceeding at such a rapid rate, the patient is hyperventilated with 100% oxygen to provide as much oxygen as possible and to eliminate carbon dioxide. Metabolic acidosis invariably occurs, so at least 44 meq of sodium bicarbonate is given and arterial blood gases are obtained as soon as possible to guide further such therapy. Renal shutdown may occur secondary to myoglobinuria, so mannitol and/or furosemide are given to help maintain urine output.

Specific measures include the administration of procaine or procainamide in doses up to 10 mg/kg (23, 24). There are as yet no reports regarding the value of dantrolene sodium in humans. Chlorpromazine may be administered to reduce shivering and produce peripheral vasodilation. Hydrocortisone in pharmacologic doses may also be of value. One instance has been reported in which cardiopulmonary bypass was successful in treating malignant hyperpyrexia (25).

If cooling is successful, it should be stopped when the body temperature has fallen to approximately 38°C in order to prevent a rebound hypothermia.

Before instituting these radical measures, one should be reasonably certain that the malignant hyperpyrexic syndrome is present. Other reasons for increases in body temperature may be existent, such as excessive draping in a warm operating room, thyrotoxicosis, or septicemia (26). However, once this syndrome has been experienced, it is never to be forgotten.

Finally, the question arises as to how one should anesthetize a patient who is known to be susceptible to the syndrome or in whom one has a high index of suspicion. Certainly, succinylcholine or potent inhalation anesthetics are to be avoided. It would appear that spinal, epidural, and regional anesthesias can be safely used, although lidocaine should be avoided as a local anesthetic because it may contribute to the development of the syndrome (27). If a general anesthetic is indicated, nitrous oxide and oxygen, supplemented by narcotics such as meperidine or by Innovar®, have been used successfully in susceptible patients. It is likely that the hypnotic thiopental and the tranquilizer diazepam can also be employed safely. If muscle relaxation is required, hyperventilation of the patient may be helpful. Until recently it was believed that D-tubocurarine could be used with impunity, but now some doubt has been cast on its safety (28).

CONCLUSION

Malignant hyperpyrexia is a genetically related syndrome that cannot be predicted reliably in advance of administration of anesthesia except by a strong family history. The definitive etiology is unknown, although triggering agents that release calcium from the calcium-storing sarcoplasmic membrane of the muscle cell are highly suspect. As soon as the syndrome is diagnosed, therapy must be prompt, vigorous, and carried out with the same urgency as a cardiac arrest. Specific therapy with dantrolene sodium may prove to be an answer to this serious problem in anesthetic practice.

Literature Cited

1. Saidman, L. J., Havard, E. S., Eger, E. I. 1964. Hyperthermia during anesthesia. *J. Am. Med. Assoc.* 190:1029–32
2. Stephen, C. R. 1967. Fulminant hyperthermia during anesthesia and surgery. *J. Am. Med. Assoc.* 202:178–82
3. Wilson, R. D., Dent, T. E., Traber, D. L., McCoy, N. R., Allen, C. R. 1967. *J. Am. Med. Assoc.* 202:183–86
4. Gordon, R. A. 1966. Malignant hyperpyrexia during general anaesthesia. *Can. Anaesth. Soc. J.* 13:415–19
5. Gordon, R. A., Britt, B. A., Kalow, W., eds. 1973. *International Symposium on Malignant Hyperthermia.* Springfield, Ill: Thomas. 473 pp.
6. Denborough, M. A., Forster, J. F. A., Lovell, R. R. H., Mapleson, P. A., Villiers, J. D. 1962. Anaesthetic deaths in a family. *Br. J. Anaesth.* 34:395–96
7. Britt, B. A., Locher, W. G., Kalow, W. 1969. Hereditary aspects of malignant hyperthermia. *Can. Anaesth. Soc. J.* 16:89–98
8. Kalow, W., Britt, B. A. 1973. Inheritance of malignant hyperthermia. See Ref. 5, pp. 67–76.
9. Kalow, W., Britt, B. A., Terreau, M. E., Haist, C. 1970. Metabolic error of muscle metabolism after recovery from malignant hyperthermia. *Lancet* 2:895–98
10. Relten, J. E. S., Britt, B. A., Steward, D. J. 1973. Malignant hyperpyrexia. *Br. J. Anaesth.* 45:269–75
11. Britt, B. A., Kalow, W., Gordon, A., Humphrey, J. G., Rewcastle, N. B. 1973. Malignant hyperthermia: An investigation of five patients. *Can. Anaesth. Soc. J.* 20:431–67
12. Hall, L. W., Woolf, N., Bradley, J. W. P., Jolly, D. W. 1966. Unusual reactions to suxamethonium chloride. *Br. Med. J.* 2:1305
13. Sybesma, W., Eikelenboom, G. 1969. Malignant hyperthermia syndrome in pigs. *Neth. J. Vet. Sci.* 2:155–60
14. Williams, C. H., Houchins, C., Shanklin, M. D. 1975. Pigs susceptible to energy metabolism in the fulminant hyperthermia stress syndrome. *Br. Med. J.* 3:411–13
15. Harrison, G. G. 1971. Anaesthetic-induced malignant hyperpyrexia: A suggested method of treatment. *Br. Med. J.* 3:454–56
16. Harrison, G. G. 1975. Control of the malignant hyperpyrexic syndrome in MHS swine by dantrolene sodium. *Br. J. Anaesth.* 47:62–65
17. Anderson, I. L., Jones, E. W. 1976. Porcine malignant hyperthermia: Effect of dantrolene sodium on in-vitro halothane-induced contraction of susceptible muscle. *Anesthesiology* 44:57–61
18. Gronent, G. A., Milde, J. H. Theye, R. A. 1976. Dantrolene in porcine malignant hyperthermia. *Anesthesiology* 44:488–95

19. Owen, G., Kerry, R. J. 1974. Anaesthesia during raised creatine phosphokinase activity. *Br. Med. J.* 4:75–76
20. Ellis, F. R., Clarke, I. M. C., Modgill, M., Currie, S., Harriman, D. G. F. 1975. Evaluation of creatinine phosphokinase in screening patients for malignant hyperpyrexia. *Br. Med. J.* 3:511–13
21. Britt, B. A., Endrenyi, L., Peters, P. L., Kwong, F. H.-F., Kadijevic, L. 1976. Screening of malignant hyperthermia susceptible families by creatine phosphokinase measurement and other clinical investigations. *Can. Anaesth. Soc. J.* 23:263–84
22. Dickey, W. T., Ahlgren, E. W., Stephen, C. R. 1970. Body temperature monitoring via the tympanic membrane. *Surgery* 67:981–84
23. Beldavs, J., Small, V., Cooper, D. A., Britt, B. A. 1971. Postoperative malignant hyperthermia: A case report. *Can. Anaesth. Soc. J.* 18:202–12
24. Noble, W. H., McKee, D., Gates, B. 1973. Malignant hyperthermia with rigidity successfully treated with procainamide. *Anesthesiology* 39:450–51
25. Ryan, J. F., Donlon, J. V., Malt, R. A. 1974. Cardiopulmonary bypass in the treatment of malignant hyperthermia. *N. Engl. J. Med.* 290:1121–22
26. Schweizer, O., Howland, W. S., Ryan, G. M., Goldines, P. L. 1971. Hyperpyrexia in the operative and immediate postoperative period. *Anesth. Analg. Cleveland* 50:906–14
27. Bianchi, C. P., Bolton, T. C. 1967. Action of local anesthetics on coupling systems in muscle. *J. Pharmacol. Exp. Ther.* 157:388
28. Britt, B. A., Webb, G. E., LeDuc, C. 1974. Malignant hyperthermia induced by curare. *Can. Anaesth. Soc. J.* 21:371–75

Ann. Rev. Med. 1977. 28:159-66
Copyright © 1977 by Annual Reviews Inc. All rights reserved

DIAGNOSIS OF PULMONARY EMBOLISM[1]

❖7233

G.V.R.K. Sharma, M.D., Donald E. Tow, M.D., Alfred F. Parisi, M.D., and Arthur A. Sasahara, M.D.
Medical, Nuclear Medicine and Research Services, Veterans Administration Hospital, West Roxbury, Massachusetts; and the Departments of Medicine and Nuclear Medicine, Peter Bent Brigham Hospital and Harvard Medical School, Boston, Massachusetts 02114

INTRODUCTION

Although there is an increasing recognition of pulmonary embolism (PE) as a major contributory factor to the mortality and morbidity of hospitalized patients, the disease continues to present difficult diagnostic problems. Recent clinical studies employing perfusion lung scan and selective pulmonary angiogram as diagnostic screening tests indicate that fatal PE occurs in about 5 patients per 1000 inpatients (1) and nonfatal PE in 20 per 1000 inpatients (2). On an annual basis, this incidence rate yields a total of approximately 140,000 fatal and 560,000 nonfatal pulmonary-embolism cases each year. While the mortality of untreated PE is high, ranging from 18% (3) to 38% (4), that of treated PE is relatively low—8% (5). It is therefore important that the clinician be aware of the implication of the above observations and institute appropriate measures for early diagnosis and treatment in the suspected patient. Diagnosis, however, may often be difficult, and at times, impossible. The diagnosis of PE can be made at the clinical level if a patient with no prior heart or lung disease presents with a history of thrombophlebitis, sudden onset of dyspnea, pleuritic pain, hemoptysis, and has minimal findings on physical examination and a "normal-looking" plain chest X ray. The clinical diagnosis may be confirmed by documentation of arterial hypoxemia and characteristic perfusion defect(s) on lung scanning. Such a presentation is so "typical" that further confirmation by selective pulmonary angiography may even be considered unnecessary. Unfortunately, however, the manifestations of pulmonary embolic disease are rarely so typical and the diagnosis rarely so easy. In the great majority of patients, the disease presents in an atypical fashion with few symptoms and fewer signs. Diagnostic difficulties are often compounded by the association of other disease processes, most notably chronic obstructive pulmonary disease (chronic bronchitis and/or emphysema) and congestive heart failure (6). Disease states like pneumonia, atelectasis, and acute myocardial infarction, which share many of the clinical features of PE, may confuse the issue. Most of the laboratory tests currently employed are nonspecific and insensitive, and render the diagnostic dilemma even more formidable. It is for this reason that in the following account an effort is made to present the factors predisposing to PE, its varied clinical picture, the available laboratory tests, and a rational diagnostic approach.

---

[1] With the technical assistance of Rick Armenia, A.B.

CLINICAL PRESENTATION

History

When the presentation is typical, the diagnostic possibility can be readily suspected. As has been pointed out earlier, however, the classical presentation of PE is infrequent. It is therefore helpful to consider the clinical setting in which the problem presents itself, since the majority of patients with PE develop their disease in a hospital during confinement for another illness. Certain disease states have been recognized as high-risk conditions for development of venous thromboembolism, bed rest and immobilization being common to these conditions. They include heart disease (especially with congestive heart failure), trauma (especially fractures of the hip), the postoperative state, chronic obstructive pulmonary disease, prior PE, and malignancy.

Other predisposing conditions include advanced age, obesity, pregnancy, and ingestion of oral contraceptive drugs. Presence of any of these predisposing factors should alert the physician to the possibility of PE, particularly in the presence of appropriate cardiopulmonary symptoms. History of preceding thrombophlebitis is a helpful pointer, but often is not forthcoming.

Symptoms and Signs

Whereas most patients without prior cardiopulmonary disease have some clinical clue suggestive of pulmonary embolism, patients with a background of chronic obstructive pulmonary disease or congestive heart failure have only nonspecific symptoms and signs. It is indeed unusual for patients with PE to present with the complete clinical picture of apprehension, dyspnea, pleuritic pain, hemoptysis, and pleural friction rub. Often, tachypnea, cough, and an accentuated pulmonic component of the second heart sound may be the only clinical findings.

In a survey carried out by us in 90 patients with angiographically proven PE, pleuritic pain and hemoptysis occurred in only 50% of the patients who had prior cardiopulmonary disease. Other symptoms and signs may be equally unrewarding. Dyspnea could be due to a variety of cardiopulmonary conditions. Fever, tachypnea, and tachycardia are not common, and if present are of no diagnostic significance in patients with chronic obstructive pulmonary disease or heart failure. Such patients are frequently likely to develop respiratory infections that may present with any of the above symptoms. However, tachypnea and tachycardia, if present, are frequently out of proportion to the clinically estimated degree of disability in patients with PE. Fever, when present, is of a moderate degree (up to 101°F); higher temperatures are generally observed in patients with extensive pulmonary infarction or in whom secondary pneumonitis has developed distal to the embolus. Generalized bronchospasm and wheezing may sometimes be seen in PE and mistaken for bronchial asthma. Signs of acute cor pulmonale such as cyanosis, elevated jugular-venous pressure, and S3 gallop with or without shock may be indicative of massive PE, but may also be secondary to chronic obstructive pulmonary disease with chronic cor pulmonale. The picture of sudden onset of chest pain with dyspnea and cardiac failure may sometimes be mistaken for myocardial infarction.

In a recent report from the National Heart and Lung Institute Urokinase Pulmonary Embolism Trial (5), an attempt was made to determine whether the massivity of the embolic process, as estimated by the perfusion lung scan and selective pulmonary angiogram, had any influence on the presenting symptoms and signs of the patients entered into the study. Severity of the embolic process was categorized as massive (more than 40% pulmonary vascular obstruction) or submassive (less than 40% obstruction). Only pleural pain was found to be significantly more frequent in the submassive group, whereas syncope occurred more frequently in those with massive embolism. The occurrence of other symptoms such as dyspnea, apprehension, cough, hemoptysis, and sweating was not influenced by the severity of the embolic occlusion. Similarly analyzed, a loud pulmonary second sound, S3 or S4 gallop sounds, and cyanosis were found more frequently in patients with massive embolism than in those with submassive occlusion. The appearance of these symptoms and signs may be helpful to the clinician, on a probability basis, in determining whether the process is a significant one or not. Massivity of the embolic process, however, did not influence the occurrence of rales, thrombophlebitis, sweating, tachypnea, tachycardia, or fever.

DIAGNOSTIC STUDIES

Laboratory Studies

Although a number of routine laboratory tests have been proposed to facilitate the diagnosis of PE, none is specific or sensitive. The white-cell count rarely exceeds 15,000 per cubic millimeter in PE, and in many instances is less than 10,000, with little shift to the left on the differential count. Initial reports (7) stressing the value of the serum enzymes, glutamic oxalacetic transaminase (GOT), lactic dehydrogenase (LDH), and serum bilirubin in establishing the diagnosis of PE have not been borne out by subsequent studies (8). Serum-enzyme elevations may be modest and evanescent, and/or secondary to nonembolic causes (e.g. congestive heart failure, pneumonia). Their absence does not militate against the diagnosis of PE (9). It is currently accepted that these biochemical studies are of little diagnostic or therapeutic significance.

When pleural effusion is present, a study of pleural fluid might be helpful. A hemorrhagic or serosanguinous effusion raises the suspicion of pulmonary infarction, but many embolic effusions can be serous (10).

Systemic arterial hypoxemia has been observed to be a consistent finding in PE, irrespective of the severity of embolic occlusion. In recent years, measurement of arterial oxygen tension has enjoyed extensive use in the diagnosis of pulmonary embolism. The test, however, has limitations in its clinical applicability and should be modified before therapeutic decisions can be based on it. A major limitation is that it cannot be applied to patients with underlying cardiopulmonary disease, who already have a lowered oxygen tension. Thus its greatest value is in excluding pulmonary embolism in young patients who have no prior history or findings of cardiopulmonary disease. In the original report of Szucs et al, patients with proven pulmonary embolism had an arterial oxygen tension of less than 80 mm Hg (11). It was therefore suggested that in patients with suspected pulmonary embolism who had an arterial Po_2 of greater than 80 mm Hg, the disease could be excluded. In the recently published Urokinase Pulmonary Embolism Trial results, however, 11.5% of the patients with proven pulmonary emoblism had an arterial Po_2 of over 80 mm Hg (5). It is recommended, therefore, that the level of arterial Po_2 by which pulmonary embolism can be excluded be set at 90 mm Hg, since almost all of the patients with pulmonary embolism will have a tension below this value. In patients with prior cardiopulmonary disease who already have a low PO_2, assuming the patient's previous ("baseline") value for arterial Po_2 is known, a further reduction with the appropriate clinical picture may point to pulmonary embolism.

Plain Chest X Ray

Most of the radiographic changes in pulmonary embolism are subtle, if not nonspecific, and therefore need careful scrutiny. The complete radiographic picture of PE described by Fleischner (12) consists of high position and diminished excursion of the diaphragm, basal atelectasis, often platelike, bilateral small pleural effusions or a large unilateral effusion, diminished vasculature on the embolized side and increased vascular markings on the unaffected side, dilatation of hilar arteries, dilated azygous vein and superior vena cava, and right-ventricular enlargement. The complete picture, however, is the exception. A recent analysis made in the multicenter trial of thrombolytic agents sponsored by the National Heart and Lung Institute (5) shows that the two findings of note are the presence of a consolidation and elevation of the diaphragm on the affected side. Even these radiographic abnormalities, although common in previously normal individuals, lose their specificity in patients with prior cardiopulmonary disease.

Electrocardiogram

The electrocardiogram usually shows little alteration in the vast majority of patients with PE. When changes do occur, they are nonspecific. The exception to this general rule is the occasional patient with acute massive PE with the characteristic S1-Q3-T3 pattern with a rightward shift of the frontal QRS axis. The more common electrocardiographic changes are T-wave abnormalities in the precordial leads and sinus tachycardia. Other electrocardiographic changes observed are rhythm disturbances (supraventricular and ventricular arrhythmias, particularly paroxysmal atrial fibrillation), varying degrees of heart block, P pulmonale, and ST changes in lateral chest leads (13). None of the above is specific for PE. The electrocardiogram, however, may be useful in excluding the possibility of acute myocardial infarction when the clinical presentation is indistinguishable.

Perfusion Lung Scan

The perfusion lung scan, because of its relatively noninvasive nature, has become the central screening procedure in patients with suspected pulmonary embolism. When four views are obtained, a normal lung scan virtually excludes clinically detectable pulmonary embolism. When the lung scan is abnormal, the likelihood of pulmonary embolism depends upon the nature of the scan defects observed. Lesions deemed to have a high probability of having been caused by pulmonary embolism are multiple and wedge-shaped or concave, according to the scan projection, and follow a segmental vascular distribution. Lesions that are classified as having a low or medium probability of having been caused by pulmonary embolism are nonsegmental and nonvascular in distribution. When the history is compatible with pulmonary embolism, the perfusion lung scan shows high-probability lesions, and the patient does not have underlying cardiopulmonary disease, treatment with intravenous heparin may be begun without further confirmation by pulmonary angiography. If, however, a nonembolic cause (e.g. pneumonia) of perfusion defect cannot be excluded, a ventilation scan, if available, may be performed. This can provide additional diagnostic information by detecting areas with loss of perfusion and retention of ventilation, characteristic of PE. Unfortunately, the majority of patients who suffer pulmonary embolism are older and generally have underlying heart or lung disease; in such instances differentiation between embolic and nonembolic scan defects cannot be made with certainty. It is therefore recommended that confirmatory pulmonary angiography be carried out in these patients.

Selective Pulmonary Angiography

Selective pulmonary angiography continues to be the only definitive means of confirming the diagnosis of PE during life, the characteristic findings being vessel cutoff and intravascular filling defect (5). Other findings, such as retarded flow, pruning, and tapered vessel caliber, are nonspecific and notoriously common in patients with chronic obstructive pulmonary disease and congestive heart failure. For this reason, in addition to the standard main pulmonary-artery injection, selective and subselective injections in the oblique views are obtained. These views bring into focus suspicious areas by improving opacification of vasculature and obtaining clearer separation of smaller vessels that are obscured in the anteroposterior projection. The angiographic study also permits hemodynamic measurements that are valuable in assessing the impact of the embolic occlusion on the cardiovascular system. Information obtained about the pulmonary arterial, capillary wedge, and right-sided filling pressures is helpful both in diagnosis and therapy. The angiographic procedure also permits evaluation of the patency of the inferior vena cava by appropriate positioning of the catheter and obtaining of an inferior vena cavagram. Information thus obtained may be useful if a caval interruption is considered.

There are limitations, however, even for pulmonary angiography. Vessels 2 mm or less cannot be visualized with sufficient clarity. Small emboli may therefore be missed. Since conventional pulmonary angiography attempts to portray the vasculature of a three-dimensional organ in a single plane, changes in the caliber and direction of vessels are difficult to interpret. Despite the limitations, pulmonary angiography yields definitive diagnostic information in the majority of suspected patients.

Assessment of Deep Veins

Since many institutions do not have angiographic facilities or a trained team necessary to perform safe studies of good quality, other, less invasive means are now being investigated as an aid to diagnosis.

Our own efforts have centered around the problem of deep-vein thrombosis and its causal relationship to pulmonary embolism. As mentioned earlier, Sevitt & Gallagher (14) noted a strong correlation at postmortem, a finding followed up some years later by Kakkar (15), using the radioactive-fibrinogen test. Since radioactive fibrinogen is not yet freely available in the United States, we chose to investigate the usefulness of the electrical-impedance plethysmographic technique, originally described by Wheeler et al (16). Noninvasive and requiring little expensive equipment, the procedure has been very useful in detecting deep-vein thrombosis. The correlation with constrast-ascending venography has been excellent. In our series of patients with pulmonary embolism, documented by pulmonary angiography, 95% had an abnormal impedance test, indicating deep-vein thrombosis (17). None of these patients had evidence or a history of recent major pelvic surgery, pelvic disease, or pelvic manipulation. In addition, from further studies it was noted that in those patients with suspected pulmonary embolism, if the impedance phlebogram was abnormal, indicating deep-vein thrombosis, 90% had a positive pulmonary angiogram whereas if the impedance phlebogram was normal, indicating patency of the deep venous system, 90% of the patients had no intravascular filling defect or arterial vessel cutoff.

The results of these studies are very promising, for they indicate the same strong relationship between deep-vein thrombosis and pulmonary embolism that was previously determined by Sevitt & Gallagher and by Kakkar. If such data regarding the inclusion or exclusion of pulmonary embolism, depending upon the assessment of the deep veins are further confirmed, invasive procedures such as pulmonary angiography may not be required in the majority of patients with suspected pulmonary embolism. A suggested diagnostic schema is illustrated in Figure 1.

Diagnostic Approach

Having recognized the possibility of PE in a given clinical setting, the physician is confronted with the task of proper investigation and prompt therapy. In spite of inadequate diagnostic sensitivity and specificity inherent in most of the tests, a careful history and physical examination with intelligent selection and interpreta-

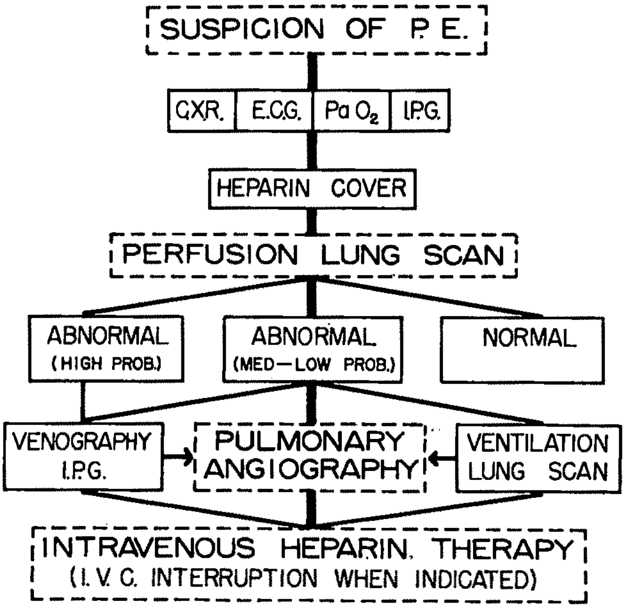


Figure 1 Suggested diagnostic schema for pulmonary embolism

tion of laboratory results allow discrimination between the various possibilities that constitute the differential diagnosis of PE. If the patient has only minimal symptoms and signs, there is adequate time for a complete laboratory workup; which is helpful in differentiating between the various diagnostic possibilities. In addition to the routine evaluation, which includes plain X ray of the chest and electrocardiogram, multiple-view lung scanning is an important aid as a screening test. Where facilities are available for performing impedance phlebography, detection of deep-venous thrombosis of the legs strengthens the suspicion of coexistent PE. Ventilation scanning appears to add specificity to perfusion scanning.

If the subject is a young individual without prior cardiopulmonary disease and the lung scan reveals characteristic perfusion defects, treatment with heparin may be started without pulmonary angiography. If, on the other hand, preexisting cardiopulmonary disease masks the clinical picture and the diagnosis is less certain, selective pulmonary angiography should be performed to definitively document or exclude the diagnosis of PE.

In the clinical setting of marked dyspnea, air hunger, anxiety, severe right-sided failure, and shock, emergency angiography is indicated, bypassing lung scanning and other less definitive investigative procedures, if a thoracic surgical team capable of performing pulmonary embolectomy is available. Among other tests, only an electrocardiogram may be important, to exclude the possibility of massive myocardial infarction, which may mimic the clinical picture. A single injection of 10,000 units of intravenous heparin is recommended to "cover" the patient before any procedure is done. If the patient improves with heparin and supportive measures, surgery may not be necessary. If, on the other hand, the condition deteriorates, with development of refractory shock or cardiopulmonary arrest, embolectomy with inferior vena caval interruption should be performed.

ACKNOWLEDGMENT

The authors wish to acknowledge the secretarial help of Miss Mary E. Stillman.

Literature Cited

1. Hume, M., Sevitt, S., Thomas, D. 1970. *Venous Thrombosis and Pulmonary Embolism,* p. 4. Cambridge, Mass: Harvard Univ. Press. 456 pp.
2. Sasahara, A. A. Unpublished data
3. Coon, W. W., Willis, P. W., Symous, M. J. 1969. Assessment of anticoagulant treatment of venous thromboembolism. *Ann. Surg.* 170:559–67
4. Barritt, D. W., Jordan, S. C. 1960. Anticoagulant drugs in treatment of pulmonary embolism: controlled trial. *Lancet* 1:1309–12
5. The Urokinase Pulmonary Embolism Trial: A national cooperative study. 1973. *Circulation* 47:Suppl. II, p. 70.
6. Sharma, G. V. R. K., Sasahara, A. A. 1975. Diagnosis of pulmonary embolism in patients with chronic obstructive pulmonary disease. *J. Chronic Dis.* 28:253–57
7. Wacker, W. E. C., Rosenthal, M., Snodgrass, P. J., Amador, E. 1961. A triad for the diagnosis of pulmonary embolism and infarction. *J. Am. Med. Assoc.* 178:8
8. Coodley, E. L. 1969. Enzyme profiles in the evaluation of pulmonary infarction. *J. Am. Med. Assoc.* 207:1307
9. Sobel, B. E. 1973. Serum enzymes and the diagnosis of PE. In *Pulmonary Thromboembolism.* Chicago: Year Book Med. Publ. 355 pp.
10. Griner, P. F. 1967. Bloody pleural effusion following pulmonary infarction. *J. Am. Med. Assoc.* 202:947
11. Szucs, M. M., Brooks, H. L., Grossman, W., Banas, J. S., Meister, S. G., Dexter, L., Dalen, J. E. 1971. Diagnostic sensitivity of laboratory findings in acute pulmonary embolism. *Ann. Intern. Med.* 74:161–66
12. Fleischner, F. G. 1965. Observations on the radiologic changes in pulmonary embolism. In *Pulmonary Embolic Disease,* ed. A. A. Sasahara, M. Stein. New York: Grune & Stratton. 312 pp.
13. Stein, P. D., Dalen, J. E., McIntyre, K. M., Sasahara, A. A., Wenger, N. K., Willis, P. K. III. 1975. The electrocardiogram in acute pulmonary embolism. *Prog. Cardiovasc. Dis.* 17:247–56
14. Sevitt, S., Gallagher, N. G. 1961. Venous thrombosis and pulmonary embolism. A clinico-pathologic study in injured and burned patients. *Br. J. Surg.* 48:475–89
15. Kakkar, V. V. 1972. Isotopic diagnosis of deep venous thrombosis. In *Thromboembolism: Diagnosis and Treatment,* ed. V. V. Kakkar, A. J. Jouhar. London: Churchill & Livingstone. 241 pp.
16. Wheeler, H. B., O'Donnell, J. A., Anderson, F. A., Benedict, K. 1974. Occlusive impedance phlebography. A diagnostic procedure for venous thrombosis and pulmonary embolism. *Prog. Cardiovasc. Dis.* 17:199–206
17. Sharma, G. V. R. K., O'Connell, D. J., Wheeler, H. B., Belko, J. S., Sasahara, A. A. 1976. Deep vein thrombosis: diagnostic clue to pulmonary embolism. Submitted for publication

Ann. Rev. Med. 1977. 28:167–77
Copyright © 1977 by Annual Reviews Inc. All rights reserved

VIRAL AGENTS AND DIARRHEAL ILLNESS

❖7234

Herbert L. DuPont, M.D., Benjamin L. Portnoy, M.D., and Richard H. Conklin, M.D., Ph.D.

The University of Texas Health Science Center at Houston, Medical School Program in Infectious Diseases and Clinical Microbiology, Houston, Texas 77030

Until recently, there was little evidence to implicate viruses as responsible agents for a sizable portion of the common form of endemic or epidemic nonbacterial gastroenteritis. However, many investigators were convinced of their importance and continued the search. In only a few years, small parvovirus-like particles were shown to cause epidemic diarrhea involving large pediatric and adult populations and somewhat later, rotaviruses (reovirus-like particles) were shown to be the major cause of infantile wintertime gastroenteritis throughout the world.

BACKGROUND

In 1945 Reimann et al (1) demonstrated that an acute diarrheal illness could be transmitted to volunteers through administration of bacteria-free filtrates of throat washings and stool suspensions obtained from others with illness. A few years later, Gordon et al (2) offered evidence that viral-like agents could be responsible for acute diarrheal illness. They infected volunteers with the "Marcy agent" after seven serial human passages. Homologous immunity was demonstrated after recovery from illness, yet Marcy agent–immune volunteers could be made ill with stool filtrates obtained from another outbreak (3). In 1971 and 1972 a series of volunteer experiments were performed with stool material from an adult patient with diarrhea acquired during secondary spread of illness at an elementary school in Norwalk, Ohio. The Norwalk agent was shown to produce illness in a high percentage of serologically unselected prison volunteers (4, 5). Homologous immunity was demonstrated when Norwalk material was again fed to volunteers following recovery from induced Norwalk disease. However, stool filtrates obtained from other outbreaks caused disease in the Norwalk-immune volunteers, indicating multiple antigenicity among viral agents (5). Kapikian et al (6) were able to directly visualize the Norwalk agent by immune electron microscopy and indicated that it was a 27-nm particle that morphologically resembles a parvovirus.

A major new direction for research in the area of viral etiology of gastroenteritis was presented by Bishop and her colleagues in 1973 (7), when viral particles were detected in the duodenal mucosa of 6 out of 9 infants and children with diarrhea studied in Australia. Morphologically similar particles resembling reoviruses, later termed rotaviruses, were soon identified in diarrheal stools by Flewett et al in England (8). Since then, rotavirus agents have been reported throughout the world and are generally recognized as the most common cause of pediatric wintertime diarrhea. While adenoviruses, ECHO viruses, Coxsackie viruses and coronaviruses occasionally produce diarrhea, because of the extent of rapidly accumulating information on rotaviruses most of the following discussion is devoted to research on this new agent.

ROTAVIRUSES

Rotavirus Morphology and Classification

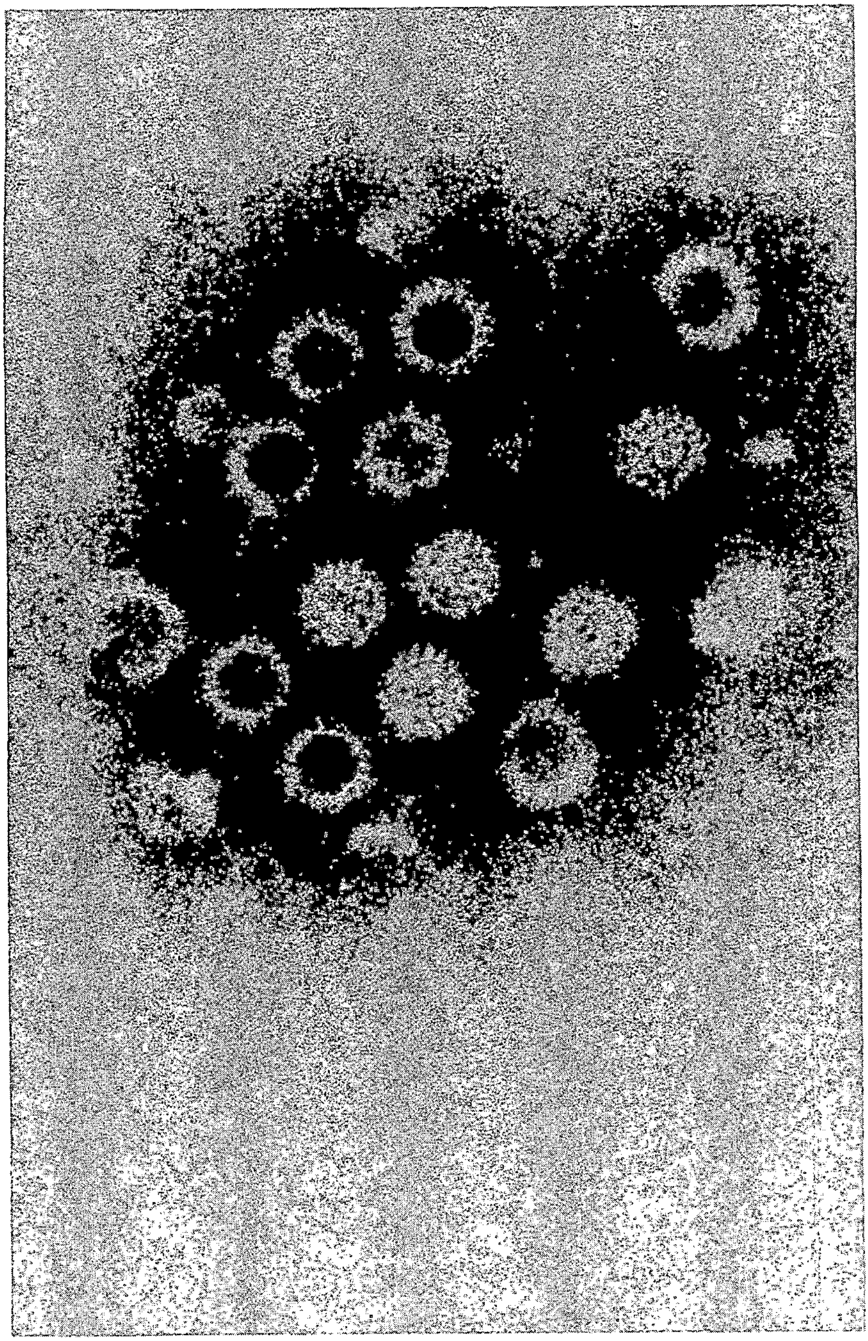
The first report of rotavirus infection was by Bishop and co-workers (7) in 1973, who tentatively called the agent an orbivirus based on size, morphology, location in the endoplasmic reticulum, and the presence of enveloped and naked particles.

Soon after the report of rotavirus from Australia, Flewett and co-workers (8) reported a similar virus in the feces of children with acute nonbacterial gastroenteritis from Great Britain. They suggested that the virus is similar to reoviruses because it has a double-capsid structure. In a subsequent report (9) they showed that the virus is morphologically identical to the Nebraska calf diarrhea virus (NCDV). In addition they noted that the viruses have a sharp, well-defined outline and capsomeres that radiate outward from a central core, like spokes of a wheel. These features are unlike those of reoviruses or orbiviruses. Flewett and co-workers (9) suggested the name rotavirus (from the Latin *rota,* a wheel), in view of the radiating capsomeres that give the virus a wheel-like appearance. Another group of investigators proposed the name duovirus (10), based on the double-capsid structure and presumed presence of double-stranded RNA. Other names that have been used include reovirus-like agent, orbivirus, and infantile gastroenteritis virus.

The virions of rotaviruses from humans, calves, foals, piglets, simians, and mice are morphologically identical (Figure 1). They are spherical particles 65–75 nm in diameter. They possess a central core 37–38 nm in diameter. Short, hollow capsomeres radiate outward from this central core. There is an additional protein layer external to this capsid surface that is sharply defined.

The human, calf, and mouse rotaviruses have been shown to be antigenically related (9–11). Immunofluorescence data (12) indicate that the calf and simian

Figure 1 Representative electron micrograph of rotaviruses in diarrheal stool from a 10-month-old Indian boy from rural Guatemala (phosphotungstic acid stain, magnified 259,000X).



rotaviruses are also immunologically cross-reactive. Neither the human nor calf rotaviruses react with antisera to reovirus types 1, 2, or 3, or to the 20 orbivirus types (9, 11).

The calf and human rotaviruses have been shown to have a buoyant density in cesium chloride of 1.36–1.37 g/ml (13, 14). Rodger and co-workers (12) report that the rotavirus polypeptide patterns, when separated on polyacrylamide gel, are more closely related to each other than to the pattern of reoviruses or orbiviruses.

The genome of the calf rotavirus has been shown by several groups to consist of double-stranded RNA. By the acridine orange method of Mayor & Diwan (15), we have shown the human rotavirus to be an RNA virus (14), which has been confirmed by labeling studies in gnotobiotic pigs infected with human rotavirus (16). A further difference of rotaviruses, when compared to orbiviruses and reoviruses, is the number of fragments in the genome (12).

The literature on this new agent is complicated by the lack of agreement on organism classification. Recently, 28 investigators working with rotaviruses reviewed the available morphologic, serologic, and physicochemical data with the hope that a tentative classification might be possible. The consensus was that a separate genus called rotavirus should be created within the family *Reoviridae* pending official classification by the Reoviridae Study Group of the International Committee of Taxonomy of Viruses (E. L. Palmer, personal communication).

Epidemiology of Rotavirus Infection

Investigators from all parts of the world have implicated rotaviruses as major causes of pediatric diarrhea. Areas where rotaviruses have been associated with diarrhea include Australia (7), New Zealand (17), Great Britain (8, 9), Canada (18), Washington DC (11), Houston (19), Atlanta (20), Guatemala (19), Mexico (unpublished data), Argentina (21), Japan (22), South Africa (23), Rhodesia (24), and India (25).

Due to the expense and insensitivity of the electron microscopy techniques currently employed in the detection of rotaviruses, it has not been possible to accurately define the modes of transmission and epidemiologic patterns of disease. Evidence thus far gathered, however, would suggest that they are responsible for approximately half of the wintertime diarrhea necessitating hospitalization in those aged 6 months to 2 years. There appears to be an inverse relationship between the seasonal patterns of rotavirus infection and shigellosis and salmonellosis (which are primarily warm-month infections). However, in a number of children with diarrhea and rotavirus agents in stool, other enteric bacterial and protozoal pathogens can be simultaneously identified (24, 26), raising the question of a possible dual infection.

While rotaviruses can occasionally be identified by electron microscopy in those without illness, their presence generally correlates well with the presence of illness. The reason for this probably relates to the large number (greater than 10^5 particles/ml) of viruses necessary to be seen by direct EM visualization. Virus concentration of this magnitude may not occur in those without infection. Rotaviruses have also been implicated in several nosocomially acquired outbreaks of diarrhea in newborn nurseries and on pediatric hospital wards, suggesting a propensity for spread within closed populations (10, 27, 28).

Most children, by two years of age, have circulating antibodies to the agent, which probably explains the sharp dropoff in occurrence of rotaviruses in older children and adults. Also, there is probably a reduced rate of illness for infants less than 6 months of age that may be explained by lack of exposure or by a protective effect of maternal antibody, which may be acquired transplacentally or through colostrum.

Clinical Manifestations of Rotavirus Infection

The incubation period of the illness is between 48 and 72 hours. The onset is generally sudden, with vomiting and watery diarrhea the most common symptoms and often associated with fever. The illness usually lasts between 2 and 8 days and in general is less severe than bacterial diarrheas. However, severe dehydration and death have been reported, though rarely, in rotavirus infection (10).

Propagation of Rotaviruses

The rotaviruses are very fastidious and do not grow in most tissue-culture systems. Calf rotavirus (NCDV) and simian rotavirus (SA–11) have been succesfully grown in tissue cultures of bovine embryonic kidney (29) and African green monkey kidney (30), and recently there have been several reports of human rotavirus infecting tissue cultures (31–34). While there has been demonstrable infection of the cells, there has not been reisolation of the virus from the cultures.

In addition to cell cultures, several other systems have been utilized to propagate the rotaviruses. Both murine and human rotaviruses have been grown in intestinal epithelial organ cultures (35, 36). It has also been possible to propagate the human rotaviruses in gnotobiotic (37, 16) piglets, conventional piglets (16), and colostrum-deprived rhesus monkeys (38). In preliminary experiments we have been able to induce diarrhea in 16 of 19 suckling rabbits with human rotavirus administered via nasogastric tube. We were unable to visualize the virus in intestinal homogenates from rabbits in the ninth day of illness, but this may be too late in the course of disease. The development of an inexpensive small-animal model should be valuable in evaluating the possibility of immunoprophylaxis and in studying the pathophysiology of this disease.

Identification of Human Rotaviruses in Patients with Illness

Since the human rotavirus has not been grown in routinely used tissue-culture cell lines, criteria for determining infection by this virus includes demonstration of the agent in a clinical specimen and/or proof of a serologic response to infection. Investigators have used a variety of techniques to diagnose rotavirus infection.

DIRECT VISUALIZATION OF THE ROTAVIRUS BY ELECTRON MICROSCOPY

The first report of rotavirus enteritis (7) demonstrated positively stained virus in duodenal biopsy specimens of children who presented with acute diarrhea. Shortly after this report, the same virus was seen in fecal extracts that were negatively stained for electron microscopy. Flewett et al (8) differentially centrifuged 20% suspensions of diarrheal stools for electron microscopic examination. Bishop et al (39) described a differential centrifugation method in which the virus is precipitated with polyethylene glycol 6000 prior to examination in the electron microscope.

The simplest electron microscopy procedure was reported by Middleton et al (18). Stool and distilled water are mixed on a microscope slide, and a drop of the suspension is placed on an electron microscope grid, stained with 1.6% phosphotungstic acid, and directly examined. The sensitivity of this technique compared to the more complex methods of differential centrifugation has not been determined. The pseudoreplica technique was used to find rotavirus by Gomez-Baretto et al (20). In this technique a drop of stool suspension is placed on a small block of highly purified agar. The fluid portion of the suspension is absorbed by the agar, and the solid particles remain on the agar surface. The particles are transferred to a Formvar film, stained with phosphotungstic acid, and examined under the electron microscope. The method is simple, rapid, and inexpensive. We have compared the pseudoreplica technique to the more time-consuming method of differential centrifugation (40) and the sensitivity of the two techniques was the same.

Electron microscopy by a variety of techniques provides a picture of rotaviruses directly from a diarrheal stool. However, other viruses in the *Reoviridae* family are similar or identical in morphology to rotaviruses, which may on occasion present problems in diagnosis.

IMMUNE ELECTRON MICROSCOPY AND IMMUNOFLUORESCENCE Immune electron microscopy was utilized by Kapikian et al (11) to detect rotaviruses. A 2% stool suspension was filtered through a 1200-nm and a 450-nm pore-size filter. The filtrate was incubated with a 1:5 dilution of serum from a patient who had recovered from a rotavirus infection and this mixture was centrifuged. The pellet was examined under the electron microscope and virus particles coated with antibody were seen. This procedure has not reached widespread acceptance, since the viral particles are readily seen without the addition of antisera.

Investigators have developed several immunofluorescence techniques for identifying rotaviruses in clinical specimens and in tissue cultures. Middleton et al (18) obtained necropsy specimens from patients who died with gastroenteritis. Cells from the duodenal and jejunal mucosa were scraped off and incubated with rotavirus antiserum prepared in a guinea pig and tagged with fluorescein. Florescein-conjugated anti–guinea pig globulin labeled the rotavirus antigen in the columnar mucosal cells. Davidson et al (41) examined duodenal biopsy specimens that had been shown by electron microscopy to contain virus. Paraffin-fixed sections of duodenal mucosa were incubated with convalescent human anti-rotavirus serum. The sections were then stained with fluorescein-labeled antisera to human globulins G, A, and M. Fluorescence was detected with antiserum to IgM only, and staining was present in the cytoplasm of epithelial cells, but not in cells of the lamina propria. It appears likely that rotavirus replication in the gut is limited to columnar epithelium of the duodenum and jejunum.

COUNTERIMMUNOELECTROPHORESIS Middleton et al (42) described and evaluated a method of counterimmunoelectrophoresis for the detection of rotavirus. Antiserum was prepared in guinea pigs to rotavirus that had been purified from feces. Stool specimens, diluted to half strength, were placed in the cathodal wells, which had been cut into agarose on a microscope slide, and the antiserum was placed in the anodal wells. Precipitin lines of antigen-antibody appeared after 60 minutes of electrophoresis. Thirty-five of 38 stools that contained rotavirus by electron microscopy were positive by counterimmunoelectrophoresis. Using this technique, one technician was able to process 20–30 specimens in 4½ hr. Counterimmunoelectrophoresis appears to be a sensitive, rapid, and inexpensive method of identifying rotavirus, and it should be useful in future epidemiologic studies.

Several methods have been reported for identifying the rotavirus in clinical specimens, but little has been written about the comparative selectivity and sensitivity of the available techniques. It is difficult to evaluate the epidemiological data reported, since each group of investigators has used different techniques of identifying the agent. Further studies are needed to determine the best methods for reliably identifying human rotavirus.

Serologic Methods of Determining Rotavirus Infection

A serologic response to rotavirus-associated diarrhea has been found by a variety of methods. Antibody titer rises may be demonstrated by immune electron microscopy, complement fixation, immune fluorescent antibody tests, and counterimmunoelectrophoresis. Correlation has been good between demonstration of rotavirus antibody production and detection of the virus in stool.

IMMUNE ELECTRON MICROSCOPY Flewett et al (9) found that serum obtained from children who had recovered from rotavirus infection agglutinated either a partially purified suspension of rotavirus or the related NCDV. Kapikian et al (11) tested paired sera from 4 patients with rotavirus-associated diarrhea for the presence of specific antibody. Using immune electron microscopy, they graded the amount of antibody coating the rotavirus particles. In all four cases an increase in the amount of antibody coating was observed among the specimens obtained during convalescence. Since the titer of antibody is difficult to quantitate by immune electron microscopy, this test is not an ideal method of routine serologic diagnosis of rotavirus infection.

COMPLEMENT FIXATION (CF) Using a stool filtrate from a patient with rotavirus-associated diarrhea as an antigen, Kapikian et al (11) demonstrated a fourfold titer rise to rotavirus in 9 patients in whom the virus had been found in the stool. Four patients with no virus in the stool had no rise in antibody titer. In a later paper (43), the same group reported a CF test for rotavirus that used NCDV as an antigen. Because NCDV may be grown in tissue culture, it serves as an ideal CF antigen. Rotavirus antigens must be prepared from stools, so it is difficult to obtain large quantities for purification of a CF antigen. Epidemiologic studies using a CF test with either antigen (20, 44) have shown good correlation between those patients with rotavirus in their stools and those with an antibody titer rise to the agent or NCDV.

IMMUNOFLUORESCENCE Echeverria et al (45) used NCDV in an immunofluorescent assay for antibodies against rotavirus. NCDV was grown on bovine embryonic kidney-cell monolayers on coverslips. The cells were fixed after 24 hours of growth and then stained by indirect immunofluorescence using serial dilutions of paired human sera and fluorescein-labeled goat anti-human globulin. In a later study (N. Blacklow, personal communication), there was a striking correlation between the presence of rotavirus in a patient's stool and an antibody titer rise by the immunofluorescent assay. This method appears to be a sensitive means of determining rotavirus infection.

SUMMARY AND CONCLUSIONS

Evidence has rapidly accumulated to implicate rotaviruses (reovirus-like agents) as the most important cause of endemic wintertime diarrhea in children under two years of age. These agents are capable of producing illness in all ages and have been found in nearly all parts of the world. Rotaviruses also are responsible for nosocomial outbreaks of gastroenteritis in pediatric hospital populations and newborn nurseries. While these agents have thus far eluded conventional virologic isolation techniques, they may be seen in direct visualization of diarrheal stool by electron microscopy. The technique is limited in usefulness, since viruses are generally seen only when they are present in massive numbers. This may explain the usual absence of viruses in stools of older children and adults with diarrhea, where there may be a lower density of viruses in stool. In one outbreak of diarrhea at an elementary school in Norwalk, Ohio, a parvovirus-like agent was identified by immune electron microscopy after a series of volunteer experiments. Such detailed studies are not possible in most outbreaks of diarrhea.

While the amount of information gathered in the past several years about viruses and human diarrheal disease is impressive, the surface has just been scratched. An immediate area for investigation is a better characterization of the epidemiology of rotavirus infection in children. Most published studies have dealt with hospitalized children with diarrhea. What role do these viruses play in the milder outpatient illness? What is the mode of transmission and where is the reservoir in nature? What is the relationship between viral agents and bacterial and protozoal pathogens? There is an urgent need for the development of a rapid and reliable serologic test for the detection of viral gastroenteritis. It appears that either the NCDV or human rotavirus will serve as a suitable antigen, and standardized serologic procedures should be forthcoming. Many of the major advances in this field must await development of procedures for the isolation of the various viral agents. It will then be possible to evaluate the specificity of serologic tests, determine the antigenic similarity of various viral pathogens, and investigate accurately the pathophysiology of viral gastroenteritis. Prevention or control of these infections, the ultimate goals, may then be possible.

ACKNOWLEDGMENTS

Rotavirus studies performed by us and reported in this paper were supported by grant AI 12699–01 of the National Institutes of Health. I. Bujnoch helped to prepare this manuscript.

Literature Cited

1. Reimann, H. A., Price, A. H., Hodges, J. H. 1945. The cause of epidemic diarrhea, nausea and vomiting (viral dysentery?). *Proc. Soc. Exp. Biol. Med.* 59:8–9
2. Gordon, I., Meneely, J. K. Jr., Currie, G. D., Chicoine, A. 1953. Clinical laboratory studies in experimentally-induced epidemic nonbacterial gastroenteritis. *J. Lab. Clin. Med.* 41:133–41
3. Fukumi, H., Nakaya, R., Hatta, S., Noriki, H., Yunoki, H., Akagi, K., Saito, T., Vchiyama, K., Kobari, K., Nakanishi, R. 1957. An indication as to identity between the infectious diarrhea in Japan and the afebrile infectious nonbacterial gastroenteritis by human volunteer experiments. *Jpn. J. Med. Sci. Biol.* 10:1–17
4. Dolin, R., Blacklow, N. R., DuPont, H. L., Formal, S., Buscho, R. F., Kasel, J. A., Chames, R. P., Hornick, R., Chanock, R. M. 1971. Transmission of acute non-bacterial gastroenteritis to volunteers by oral administration of stool filtrates. *J. Infect. Dis.* 123:307–12
5. Wyatt, R. G., Dolin, R., Blacklow, N. R., DuPont, H. L., Buscho, R. F., Thornhill, T. S., Kapikian, A. Z., Chanock, R. M. 1974. Comparison of three agents of acute infectious nonbacterial gastroenteritis by cross-challenge in volunteers. *J. Infect. Dis.* 129:709–14
6. Kapikian, A. Z., Wyatt, R. G., Dolin, R., Thornhill, T. S., Kalica, A. R., Chanock, R. M. 1972. Visualization by immune electron microscopy of a 27-nm particle associated with acute infectious non-bacterial gastroenteritis. *J. Virol.* 10:1075–81
7. Bishop, R. F., Davidson, G. P., Holmes, I. H., Ruck, B. J. 1973. Virus particles in epithelial cells of duodenal mucosa from children with acute non-bacterial gastroenteritis. *Lancet* 2:1281–83
8. Flewett, T. H., Bryden, A. S., Davies, H. 1973. Virus particles in gastroenteritis. *Lancet* 2:1497
9. Flewett, T. H., Bryden, A. S., Davies, H., Woode, G. N., Bridger, J. C., Derrick, J. M. 1974. Relation between viruses from acute gastroenteritis of children and newborn calves. *Lancet* 2:61–63
10. Davidson, G. P., Bishop, R. F., Townley, R. R. W., Holmes, I. H., Ruck, B. J. 1975. Importance of a new virus in acute sporadic enteritis in children. *Lancet* 1:242–46
11. Kapikian, A. Z., Kim, H. W., Wyatt, R. G., Rodriguez, W. J., Ross, S., Cline, W. L., Parrott, R. H., Chanock, R. M. 1974. Reovirus-like agent in stools: Association with infantile diarrhea and development of serologic tests. *Science* 185:1049–53
12. Rodger, S. M., Schnagl, R. D., Holmes, I. H. 1975. Biochemical and biophysical characteristics of diarrhea viruses of human and calf origin. *J. Virol.* 16:1229–35
13. Welch, A. B. 1971. Purification, morphology and partial characterization of a reovirus-like agent associated with neonatal calf diarrhea. *Can. J. Comp. Med.* 35:195–202
14. Conklin, R. H., Portnoy, B. L., Spindler, S., DuPont, H. L. 1975. Viral agents in acute childhood diarrhea from Houston and Guatemala. *Intersci. Conf. Antimicrob. Agents Chemother., 15th, Washington DC* (Abstr.)
15. Mayor, H. D., Diwan, A. R. 1961. Studies on acridine orange staining of two purified RNA viruses: Poliovirus and tobacco mosaic virus. *Virology* 14:74–82
16. Middleton, P. J., Petric, M., Szymanski, M. T. 1975. Propagation of infantile gastroenteritis virus (orbi-group) in conventional and germfree piglets. *Infect. Immun.* 12:1276–80
17. Barnes, G. 1975. Duovirus in New Zealand (letter). *Lancet* 1:1192
18. Middleton, P. J., Szymanski, M. T., Abbott, G. D., Bortolussi, R., Hamilton, J. P. 1974. Orbivirus acute gastroenteritis of infancy. *Lancet* 1:1241–44
19. Conklin, R. H., DuPont, H. L., Goldschmidt, M. C., Rodriguez, J. T. 1975. Occurrence of "viral particles" in diarrhea: Houston, Texas and Guatemala. *N. Engl. J. Med.* 292:644–45

20. Gomez-Barreto, J., Palmer, E. L., Nahmias, A. J., Hatch, M. H. 1976. Acute enteritis associated with reovirus-like agents. *J. Am. Med. Assoc.* 235:1857–60
21. Lombardi, G. H., Roseto, A. M., Stamboulian, D., Oro, J. G. B. 1975. Virus of infantile gastroenteritis in Argentina (letter). *Lancet* 2:1311
22. Konno, T., Suzuki, H., Ishida, N. 1975. Reovirus-like agent in Japanese infants with gastroenteritis (letter). *Lancet* 1:918–19
23. Schoub, B. D., Koornhof, H. J., Lecatsas, G., Prozesky, O. W., Freiman, I., Hartman, E., Kassel, H. 1975. Viruses in acute summer gastroenteritis in black infants (letter). *Lancet* 1:1093–94
24. Cruickshank, J. G., Zilberg, B., Axton, J. H. M. 1975. Virus particles and gastroenteritis in black and white children in Rhodesia. *S. Afr. Med. J.* 49:859–63
25. Holmes, I. H., Mathan, M., Bhat, P., Albert, M. J., Swaminathan, S. P., Maiya, P. P., Pereira, S. M., Baker, S. J. 1974. Orbivirus and gastroenteritis. *Lancet* 2:658–59
26. Evans, D. G., Olarte, J., DuPont, H. L., Evans, D. J. Jr., Galindo, E., Portnoy, B. L., Conklin, R. H. Enteropathogens associated with pediatric diarrhea in Mexico City. *J. Pediatr.* In press
27. Cameron, D. J. S., Bishop, R. F., Davidson, G. P., Townley, R. R. W., Holmes, I. H., Ruck, B. J. 1976. New virus associated with diarrhoea in neonates. *Med. J. Aust.* 1:85–86
28. Bishop, R. F., Hewstone, A. S., Davidson, G. P., Townley, R. R. W., Holmes, I. H., Ruck, B. J. 1976. An epidemic of diarrhoea in human neonates involving a reovirus-like agent and "enteropathogenic" serotypes of *Escherichia coli. J. Clin. Pathol.* 29:46–49
29. Malherbe, H. H., Strickland-Cholmley, M. 1967. Simian virus SA 11 and the related O agent. *Arch. Gesamte. Virusforsch.* 22:235–45
30. Fernelius, A. L., Ritchie, A. E., Classick, L. G., Norman, J. O., Mebus, C. A. 1972. Cell culture adaptation and propagation of a reovirus-like agent of calf diarrhea from a field outbreak in Nebraska. *Arch. Gesamte. Virusforch.* 37:114–30
31. Banatvala, J. E., Totterdell, B., Chrystie, I. L., Wood, G. N. 1975. In-vitro detection of human rotaviruses (letter). *Lancet* 2:821
32. Wyatt, R. G., Gill, V. W., Sereno, M. M., Kalica, A. R., VanKirk, D. H., Chanock, R. M., Kapikian, A. Z. 1976. Probable in-vitro cultivation of human reovirus-like agent of infantile diarrhea. *Lancet* 1:98–99
33. Purdham, D. R., Purdham, P. A., Evans, N., McNeish, A. S. 1975. Isolation of human rotaviruses using embryonic gut monolayers. *Lancet* 2:977
34. Albrey, M. B., Murphy, A. M. 1976. Rotavirus growth in bovine monolayers. *Lancet* 1:753
35. Rubenstein, D., Milne, R. G., Buikland, R., Tyrrell, D. A. V. 1971. The growth of the virus of epidemic diarrhea of infant mice (E.D.I.M.) in organ cultures of intestinal epithelium. *Br. J. Exp. Pathol.* 52:442–45
36. Wyatt, R. G., Kapikian, A. Z., Thornhill, T. S., Sereno, M. M., Kim, H. W., Chanock, R. M. 1974. In vitro cultivation in human fetal intestinal organ cultures of a reovirus-like agent associated with nonbacterial gastroenteritis in infants and children. *J. Infect. Dis.* 130:523–28
37. Torres-Medina, A., Wyatt, R. G., Mebus, C. A., Underdahl, N. R., Kapikian, A. Z. 1976. Diarrhea caused in gnotobiotic piglets by the reovirus-like agent of human infantile gastroenteritis. *J. Infect. Dis.* 133:22-27
38. Wyatt, R. G., Sly, D. L., London, W. T., Palmer, A. E., Kalica, A. R., Van Kirk, D. H., Chanock, R. M., Kapikian, A. Z. 1976. Induction of diarrhea in colostrum-deprived newborn rhesus monkeys with the human reovirus-like agent of infantile gastroenteritis. *Arch. Virol.* 50:17–27
39. Bishop, R. F., Davidson, G. P., Holmes, I. H., Ruck, B. J. 1974. Detection of a new virus by electron microscopy of faecal extracts from children with acute gastroenteritis. *Lancet* 1:149–51.
40. Portnoy, B. L., Conklin, R. H., Menn, M., Olarte, J., DuPont, H. L. 1976. Reliable identification of reovirus-like agent in diarrheal stools. *J. Lab. Clin. Med.* In press
41. Davidson, G. P., Goller, I., Bishop, R. F., Townley, R. R. W., Holmes, I. H., Ruck, B. J. 1975. Immunofluorescence in duodenal mucosa of children with acute enteritis due to a new virus. *J. Clin. Pathol.* 28:263–66
42. Middleton, P. J., Petric, M., Hewitt, C. M., Szymanski, M. T., Tam, J. S. 1976. Counter-immunoelectro-osmophoresis for the detection of infantile gastroen-

teritis virus (orbi-group) antigen and antibody. *J. Clin. Pathol.* 29:191–97

43. Kapikian, A. Z., Cline, W. L., Mebus, C. A., Wyatt, R. G., Kalica, A. R., James, H. D. Jr., VanKirk, D., Chanock, R. M. 1975. New complement-fixation test for the human reovirus-like agent of infantile gastroenteritis. *Lancet* 1:1056–61
44. Kapikian, A. Z., Kim, H. W., Wyatt, R. G., Cline, W. L., Arrobio, J. O., Brandt, C. D., Rodriguez, W. J., Sack, D. A., Chanock, R. M., Parrott, R. H. 1976. Human reovirus-like agent as the major pathogen associated with "winter" gastroenteritis in hospitalized infants and young children. *N. Engl. J. Med.* 294:965–72
45. Echeverria, P., Blacklow, N. R., Smith, D. H. 1975. Role of heat labile toxigenic *Escherichia coli* and reovirus-like agent in diarrhea in Boston children. *Lancet* 2:1113–16

Ann. Rev. Med. 1977. 28:179–85
Copyright © 1977 by Annual Reviews Inc. All rights reserved

PERTUSSIS IN THE ADULT

❖7235

Calvin C. Linnemann, Jr., M.D. and Jerome Nasenbeny, M.D.
Infectious Disease Division, Departments of Medicine and Pediatrics, University of Cincinnati College of Medicine; and the Children's Hospital Medical Center, Cincinnati, Ohio 45267

INTRODUCTION

> I believe . . . that pertussis in adults is often unrecognized, and this is accountable in many cases for the spread of the disease. In one year in three different families, I saw six cases of pertussis contracted from nurses who had coughs, which they thought were due to "ordinary colds." The prevalent idea among the laity, and sometimes among physicians and nurses, that adults seldom contract whooping cough is responsible for lack of quarantine.
>
> John Phillips, 1921 (1)

Over 50 years later, and several decades after the development of an effective pertussis vaccine, the "prevalent idea" that adults do not contract pertussis continues to result in the spread of disease. An example of this occurred in 1974, when pertussis developed in pediatricians and pediatric nurses at the University of Cincinnati and spread to infants in the newborn nursery before the cause of the epidemic was recognized (2). The reason for the delay in recognition was the belief that adults do not develop the disease.

Although pertussis is primarily a disease of the very young, cases have always occurred in adult contacts of children with the disease. This is reflected in some of the colloquial names for pertussis such as "grandmother's whooping cough" or "nurses' cough" (3). Widespread use of effective pertussis vaccines has contributed to the decreasing incidence of clinically recognizable pertussis in children, but recent observations have suggested that the disease is continuing to occur in adults. This report reviews the clinical and epidemiological aspects of pertussis in adults, the currently available control measures, and future prospects concerning the disease.

CLINICAL MANIFESTATIONS

The term *pertussis* has replaced "whooping cough" to indicate infection with *Bordetella pertussis.* This reflects the recognition that *B. pertussis* produces a spectrum of clinical manifestations, from mild respiratory-tract infection to severe paroxysmal coughing associated with an inspiratory whoop and occasionally bronchopneumonia. Most studies of pertussis have been confined to patients with the more severe clinical illness, primarily because of the technical difficulties in obtaining a bacteriologic diagnosis. One study of the clinical manifestations of *Bordetella* infections reported that more than half of those with pertussis had paroxysmal coughs, a third had nonparoxysmal coughs, and 1% or less had inapparent infections (4). The clinical illness may vary with increasing age, immunization, or previous infection, but this has not been carefully studied.

Pertussis in the adult may be similar to disease in the young infant. This was vividly illustrated during the hospital epidemic at the University of Cincinnati in 1974, when one of the authors had the opportunity to experience pertussis in the adult in the first person:

"I am a 29-year-old white male pediatric resident who became ill in March 1974, while rotating through the newborn nursery at the Cincinnati General Hospital. My initial symptoms were coryza, sneezing, and an irritating nonproductive cough without fever. The cough gradually increased in frequency over the first 2 weeks until I was having explosive paroxysms of coughing 6 to 10 times an hour, with each paroxysm lasting 5 to 10 seconds. I also developed hoarseness during the second week, which persisted throughout the illness. Initially, the cough occurred during the day, but in the second week, coughing occurred throughout the night and appeared to be aggravated by a recumbent position. This made sleeping impossible, and codeine did not suppress the cough. At this time, a nurse's aid in the nursery suggested that I had "whooping cough," but this was discounted as a joke. By the end of the second week, the cough was still nonproductive, but now the paroxysms were followed by an inspiratory whoop, which can only be compared to a feeling of suffocation. Even with a voluntary effort, the inspiration could not be stopped. The few seconds involved seemed much longer.

For the first time, the diagnosis of pertussis was suggested seriously, but still considered unlikely. A cough-plate culture was made, which was later overgrown by normal mouth flora, and ampicillin was started for my "bronchitis." A white blood-cell count and differential at this time and on several subsequent occasions were normal. A nasopharyngeal swab was cultured on the fifth day of ampicillin therapy and *B. pertussis* was isolated. The ampicillin was changed to erythromycin. By the third week, my cough became productive and was frequently followed by vomiting. During the fourth week, I developed left lower back pain which was treated with bed rest, local heat, and Valium. During this time the cough was not triggered by any specific events.

In the fifth week, the paroxysms began to decrease in number and severity, and for the first time I could anticipate when a paroxysm was going to begin. By the sixth week, I was able to return to work, seeing patients in the outpatient clinic. Several days after this, I developed coryza and the paroxysmal coughing returned. That evening I had a temperature of 104°F and became tachypneic. After one burst of paroxysms which lasted 15 to 20 minutes, I had great difficulty breathing. A culture and fluorescent-antibody stain for *B. pertussis* at that time were negative, but influenza A was recovered from a pharyngeal swab. This exacerbation decreased over the next few days, but a chronic productive cough continued into the ninth week.

Although my medical records from childhood are not available, my mother believes that I received the customary 'baby shots.' "

This illustrates the potential for *B. pertussis* to produce classical whooping cough in adults. The only exception was the absence of the lymphocytosis that is characteristic of pertussis in children (5). Mannerstedt studied 29 adults with pertussis, six of whom were confirmed by culture, and reported that sequential white blood-cell counts were normal (6). However, he did note that in some patients the absolute lymphocyte count decreased with recovery. Another study of vaccinated children stated that the white blood-cell count was not useful in diagnosis, suggesting that partial immunity might interfere with the development of lymphocytosis (7).

Severe pertussis appears to be unusual in adults, although the clinical spectrum of infection in the older age groups has not been clearly defined. Mannerstedt's study of pertussis in the adult was done in 1934, before the introduction of effective vaccines (6). His patients were adults seen in juvenile pertussis clinics, apparently accompanying sick children, and were selected because they had chronic coughs. He did not study asymptomatic adults or those with minor respiratory symptoms to determine whether they were infected. Therefore his observations are limited to those whose illness progressed to chronic coughing. He reported that the symptoms of his patients persisted from 24–80 days, with an average duration of 6 weeks. In 23 of his 29 patients the cough was paroxysmal, and 27 had productive coughs. He noted the variable presence of whooping, vomiting, and choking, but did not indicate the frequency with which these occurred.

Smith reported a school epidemic of pertussis in teenage boys (8). Three out of twenty boys with positive cultures had inspiratory whoops. There was a wide variety of clinical manifestations, ranging from a "series of single coughs," " 'tickly' cough," and "slight cough on waking" to those with frequent paroxysms of coughing and whooping. He did not define the duration of illness, but commented that some continued to cough for 2 months. Luttinger described a group of adults who were the apparent source of infection for children (9). They had only a "little cough" but this lasted for 5–6 weeks. In our experience during the epidemic in Cincinnati, only one of eight adults with confirmed pertussis developed severe disease with "whooping", but all were sick for over 2 weeks (2). Therefore, although the symptoms in older children and adults may be less severe than in younger children, the illness in many cases may last for a prolonged period of time. This is an interesting observation, in view of a recent survey of pertussis in England that indicated that the severity of illness and complications in children were age-related, but the length of illness showed little difference among various age groups (10).

EPIDEMIOLOGICAL ASPECTS

Prior to the introduction of effective pertussis vaccines, most people were infected at an early age. The maximum incidence occurred in the preschool-age group (11).

Natural infection provides long-lasting immunity, so clinically recognizable infections were rare in the older age groups. Reinfections did occur, as indicated by the appearance of mild disease in parents of infected children, but the extent of reinfections in older age groups and their role in maintaining immunity to pertussis were not defined. Children with clinical pertussis were the primary reservoir of infection, since a true carrier state does not exist (12). Adults with pertussis did serve as a source of infection for some children. In 1916, Luttinger described a "Pertussis Pete," a young man with a chronic cough due to pertussis who infected a number of children, and suggested that adults with pertussis may play a role in the dissemination of disease (9). He also reviewed several German studies, one of which reported a doctor acquiring pertussis in the hospital and transmitting it to his family (13). Recent reports of hospital-associated epidemics have shown that this pattern of transmission by adults still occurs, but do not indicate how frequently it occurs (2, 14). In large series of pertussis patients, 2% or less of cases were adults (9).

The widespread use of pertussis vaccine in the United States in the 1940s changed the epidemiology of the disease. Mortality rates were decreasing prior to this time, but the incidence of disease was not (15). The use of pertussis vaccines was associated with a decrease in the number of reported cases and a relative shift to an older age of attack. Ipsen & Bowen reported a decreasing incidence of pertussis in children under 10 years of age in Massachusetts beginning in the 1940s, but not in children in the 10-to-14-year age group (16). In 1958 Eldering & Kendrick reported that more of the cases in Grand Rapids, Michigan were occurring in children 10 years of age and older, but that there was not an increase in the attack rate in this age group (17, 18). The proportion of total cases was higher in the older age group because of the decrease in attack rates in the younger age groups. Similar changes in the age distribution in children have been reported in European countries (19). A serologic survey in Czechoslovakia after the introduction of pertussis vaccine showed that antibody was present more often in young children and older adults than in young adults (19). Presumably, vaccine produced antibody in the children and natural infection accounted for antibody in older adults.

None of these epidemiologic studies were specifically concerned with pertussis in the adult. The changing age distribution was attributed to the protection of younger children by vaccine, while older, unvaccinated children continued to develop disease. Subsequent studies demonstrated that older, vaccinated individuals also developed pertussis (20). Lambert reported an attack rate of 95% in individuals who had been vaccinated 12 or more years prior to exposure to pertussis. This loss of vaccine-induced immunity suggests the possibility that the incidence of disease might increase in previously vaccinated adults, but this has not been demonstrated. The proportion of cases occurring in adults in some recent epidemics has ranged from 20 to 33% (20–22). In Lambert's study, 39 of 186 (21%) of those 20 years of age or older developed pertussis; in a small epidemic in Oregon, 6 of 16 (38%); and in a family study in Los Angeles, California in 1974–1975, 11% (20, 22, 23). The proportion of cases and the attack rates in adults seem higher than would be expected from the pre-vaccine studies. However, the recent reports are intensive epidemic investigations or family studies. It is possible that similar studies in the pre-vaccine era would have given the same results with regard to attack rates in the older age groups.

CONTROL MEASURES

The recognition that pertussis occurs in adults and that vaccine-induced immunity is lost with time raises the question of revaccinating adults. At the present time, pertussis vaccine is recommended for the routine immunization of children up to 6 years of age. The vaccine is not used in older age groups because the risk:benefit ratio is considered too high, although estimations of the risks are extremely variable (24). Reduced doses of currently available pertussis vaccines (0.2 ml) have been used in adults, with seroconversion occurring in 75% of those vaccinated (2, 25, 26). Local reactions constitute a problem that would interfere with any attempt to routinely revaccinate adults; however, in epidemic situations such as a hospital outbreak, vaccination of adults may be indicated. In one such epidemic, 885 hospital workers were vaccinated (2). The majority of vaccinees had local reactions; 10% developed fevers ($\geq$ 100°F); rashes occurred in several vaccinees, including erythema multiforme; and one man, who had a history of a febrile convulsion in childhood, developed a transient episode of memory difficulty. Appropriate screening might have eliminated the latter problem, but one must be willing to accept local reactions and occasional allergic reactions if the vaccine is used in adult populations.

Antibiotics do not affect the course of pertussis, except perhaps when started early in the infection, but they do eradicate *B. pertussis* from respiratory-tract secretions (27). Therefore antibiotics are useful in preventing further spread of the disease from infected patients. Uncontrolled observations have also indicated that erythromycin chemoprophylaxis may be effective (2, 27). Until the appropriate controlled studies are done, the use of erythromycin would seem reasonable in adults in close contact with pertussis patients or in the containment of epidemics in limited populations. In these situations, 1–1½ g of erythromycin should be given daily for 14 days. If the individual has continuing exposure to new cases of pertussis, then vaccination should be considered.

FUTURE PROSPECTS

Attempts to produce an effective vaccine that does not produce reactions have been unsuccessful (28). The protective antigen cannot be separated entirely from the antigens or toxic fractions that produce those reactions. One alternative that has been suggested for revaccination of adults is the intranasal administration of the vaccine (29). This produces a secretory immune response in the respiratory tract, but no increase in serum antibody. Evaluation of the efficacy of intranasal vaccination in the prevention of infection would be very difficult. The decrease in the recognized morbidity and mortality of pertussis has been associated with a decrease in interest in vaccine development. The increasing attention to the occurrence of pertussis in older age groups, if supported by the appropriate surveillance for *B. pertussis* in the general population, could provide the basis for renewed interest in vaccine development.

A related problem, which has not been discussed in this review, is the occurrence of another *Bordetella* infection, caused by *B. parapertussis.* Parapertussis is a milder disease than pertussis, but approximately 40% develop prolonged bronchitis (4). *B. parapertussis* is extremely widespread, as shown by antibody surveys in which the majority of adults have been infected (19, 30). The diagnosis is seldom made, because bacteriologic laboratories do not routinely culture for this organism, but recent experience confirms that parapertussis is a significant problem (31, 32). *B. parapertussis* infects adults, just as does *B. pertussis* (4). Culture surveys of adult populations may show that both pertussis and parapertussis contribute significantly to the etiology of prolonged bronchitis in adults.

Literature Cited

1. Phillips, J. 1921. Whooping-cough contracted at the time of birth, with report of two cases. *Am. J. Med. Sci.* 161:163–65
2. Linnemann, C. C. Jr., Perlstein, P. H., Ramundo, N., Minton, S. D., Englender, G. S. 1975. Use of pertussis vaccine in an epidemic involving hospital staff. *Lancet* 2:540–51
3. Madsen, T. 1937. *Lectures on the Epidemiology and Control of Syphilis, Tuberculosis, and Whooping Cough, and Other Aspects of Infectious Disease,* p. 208. Baltimore: Williams & Wilkins. 216 pp.
4. Vysoka, B. 1958. The epidemiology of pertussis and parapertussis. *J. Hyg. Epidemiol. Microbiol. Immunol.* 2:196–204
5. Lagergren, J. 1963. The white blood cell count and the erythrocyte sedimentation rate in pertussis. *Acta Paediatr. Stockholm* 52:405–9
6. Mannerstedt, G. 1934. Pertussis in adults. *J. Pediatr.* 5:596–600
7. Jaffe, V. R. 1955. Incidence of pertussis in vaccinated and unvaccinated children. *J. Pediatr.* 47:716–19
8. Smith, R. E. 1936. A review of recent work on whooping-cough. *Q. J. Med.* 5(19):307–26
9. Luttinger, P. 1916. The epidemiology of pertussis. *Am. J. Dis. Child.* 12:290–315
10. Miller, C. L., Fletcher, W. B. 1976. Severity of notified whooping cough. *Br. Med. J.* 1:117–19
11. Collins, S. D. 1929. Age incidence of the common communicable diseases of children. *Publ. Health Rep.* 44:763–826
12. Linnemann, C. C. Jr., Bass, J. W., Smith, M. H. D. 1968. The carrier state in pertussis. *Am. J. Epidemiol.* 88:422–27
13. Schwenkenbecher, A. 1921. Keuchhusten bei erwachsenen. *Med. Klin.* 17:1455–56
14. Kurt, T. L., Yeager, A. S., Guenette, S., Dunlop, S. 1972. Spread of pertussis by hospital staff. *J. Am. Med. Assoc.* 221:264–67
15. Gordon, J. E., Aycock, W. L. 1951. Whooping cough and its epidemiological anomalies. *Am. J. Med. Sci.* 222:333–61
16. Ipsen, J., Bowen, H. E. 1955. Effects of routine immunization of children with triple vaccine (diphtheria-tetanus-pertussis). *Am. J. Publ. Health* 45:312–18
17. Eldering, G. 1958. Whooping cough studies in Grand Rapids and Kent County. *J. Mich. Med. Soc.* 57:221–24
18. Kendrick, P. L. 1958. Some developments in relation to pertussis vaccination. *J. Mich. Med. Soc.* 57:693–97
19. Vysoka-Burianova, B. 1963. Contemporary problems in the epidemiology of whooping cough. *J. Hyg. Epidemiol. Microbiol. Immunol.* 7:472–81
20. Lambert, H. J. 1965. Epidemiology of a small pertussis outbreak in Kent County, Michigan. *Publ. Health Rep.* 80:365–69
21. Agate, G. H., Eldering, G. 1967. Pertussis—Michigan. *Morbid. Mortal. Rep.* 16:379
22. Goldblatt, E. L. 1968. Pertussis—Oregon. *Morbid. Mortal. Rep.* 17:6
23. Baraff, L. J., Wilkins, J., Wehrle, P. F. 1976. Epidemiology of pertussis—Los

Angeles County. *Clin. Res.* 24:183A (Abstr.)

24. Edsall, G. 1975. Present status of pertussis vaccination. *Practitioner* 215: 310–14
25. Volk, V. K., Gottshall, R. Y., Anderson, H. D., Top, F. H., Bunney, W. E., Serfling, R. E. 1964. Antibody response to booster dose of diphtheria and tetanus toxoids and pertussis vaccine. *Publ. Health Rep.* 79:424–34
26. Kendrick, P. L., Gottshall, R. Y., Anderson, H. D., Volk, V. K., Bunney, W. E., Top, F. H. 1969. Pertussis agglutinins in adults. *Publ. Health Rep.* 84:9–15
27. Bass, J. W., Klenk, E. L., Kotheimer, J. B., Linnemann, C. C., Smith, M. H. D. 1969. Antimicrobial treatment of pertussis. *J. Pediatr.* 75:768–81
28. Pittman, M. 1970. In *Infectious Agents and Host Reactions,* ed. S. Mudd, pp. 239–70. Philadelphia: Saunders. 626 pp.
29. Thomas, G. 1975. Respiratory and humoral immune response to aerosol and intramuscular pertussis vaccine. *J. Hyg. Cambridge* 74:233–37
30. Flosdorf, E. W., Bondi, A., Felton, H., McGuinness, A. C. 1942. Studies with *Hemophilus pertussis.* X. Comparative antigenic analysis of *Bacillus parapertussis* and *Hemophilus pertussis,* phase I, with consideration of clinical significance. *J. Pediatr.* 21:625–34
31. Lautrop, H. 1971. Epidemics of parapertussis. 20 years' observations in Denmark. *Lancet* 1:1195–98
32. Linnemann, C. C. Jr., Perry, E. B. 1976. *Bordetella parapertussis:* Recent experience and a review of the literature. *Am. J. Dis. Child.* In press

Ann. Rev. Med. 1977. 28:187–94
Copyright © 1977 by Annual Reviews Inc. All rights reserved

PAROXYSMAL NOCTURNAL HEMOGLOBINURIA (PNH) AS A CLONAL DISORDER

❖7236

Robert C. Hartmann, M.D. and Anita B. Arnold
Section of Hematology/Oncology, Department of Internal Medicine, College of Medicine, University of South Florida; and Veterans Administration Hospital, Tampa, Florida 33612

INTRODUCTION

Recent years have witnessed accelerated interest in the clonal nature and origin of disease. Benign disorders of clonal nature are probably the result of single mutations that may or may not be followed by essentially normal growth patterns. On the other hand, the somatic-mutation theories of cancer imply unicellular or clonal origin, which proceeds, however, through progressive clonal growth and selection to fully malignant characteristics (1). In view of recent reports of paroxysmal nocturnal hemoglobinuria (PNH) terminating in acute leukemia (2–4),[1] it seems pertinent to review the evidence for PNH as a clonal disorder.

Nowell views the natural history of tumor progression as a "microevolutionary" process, with continuing production of a multiplicity of variant cells (6). Among the variants there occurs an occasional mutant with selective growth advantage permitting it to overgrow not only normal cells in the area, but the parental tumor-cell line as well. Such processes may occur repeatedly until frankly invasive malignancy develops. This theory is basically Darwinian, with the analogy between evolution of the species and microevolution of malignancy. This is, for example, consonant with postulated progression of uterine cervical dysplasia to carcinoma in situ and finally to malignant, invasive cervical carcinoma. This theory does not exclude the influence of other factors, i.e. viruses and carcinogens, via effects on the genome.

Support for the microevolutionary theory of tumorogenesis derives in a major way from chromosome studies (6). Although not all tumors show chromosomal aberrations, as most tumors progress the chromosomal patterns tend to become (more) abnormal, parallel with the loss of specialized function of the ancestral normal cell and with improved capacity for growth and invasion (7).

---

[1] Space considerations preclude a seemingly interminable listing of individual case reports and studies. Reference is instead made to key reviews (2–5) in which individual references have been compiled and discussed, all with full recognition that the original work stems from a large number of investigators.

An important model for study of the clonal nature of disease stems from the mosaicism produced by random functional inactivation of one or the other of the two X chromosomes in female somatic cells early in embryonic life (Lyon phenomenon). Using females heterozygous for the sex-linked enzyme glucose-6-phosphate dehydrogenase (G-6-PD) types A and B, it has been shown that at least 14 human tumors are unicellular in origin (1,8), e.g. cultured tumor cells possess only one type of G-6-PD. Unicellular origin has also been demonstrated for benign processes such as hereditary X-linked sideroachrestic anemia (9) and for the repair of atherosclerotic plaques (1). On the other hand, a smaller number of tumors appear to be of multicellular origin (1). In heritable tumors of multicellular origin, it is possible that an initial or "preparatory" carcinogenic event occurred prior to X-chromosome inactivation, and a second or precipitating carcinogenic event later involves all cell types. Other speculations have been invoked to explain multicellular origin of ostensibly nonheritable tumors (1).

The study of blood disorders has yielded a rich source of information. A prerequisite for the clonal-evolution theory of tumorogenesis is the existence of a "variant commun" for the individual type of tumor that permits subsequent microevolution to fully developed malignant cell clones. The best theoretical example is the Ph^1 chromosome of chronic myeloid leukemia (CML), which may even be present in marrow preceding onset of clinically detectable disease (1). CML has been viewed as "carcinogenesis in slow motion" and variously termed as a "conditioned malignancy" or as "premalignant." The malignant progression and clonal evolution are slow until late in the course (i.e. blastic transformation), and then a large variety of cytogenetic abnormalities may develop (10).

PNH AS A CLONAL DISORDER

PNH has been associated with CML and "related" disorders (acute leukemia, myeloid metaplasia) and with aplastic anemia (2–5). Whether the inclusion of PNH with the "myeloproliferative disorders" (11) and its designation as a "malignant disorder" (1) are justified is open to question. Nonetheless, such suggestions render pertinent a review of the evidence for PNH as a clonal disorder. The salient clinical and laboratory features of PNH have been reviewed elsewhere (2–5).

The finding of low acetylcholinesterase (AChE) activity in PNH erythrocytes (RBCs) suggested an heritable disorder, but this was quickly vitiated by the negative results of many family studies. An acquired mutation was then suggested (12), and from this evolved the clonal theory for PNH (13). Variously expressed, but put simply, this implies the "development or ascendancy of a new clone of normoblasts giving rise to circulating PNH RBCs, perhaps in the setting of marrow damage" (e.g. hypoplastic anemia).

Supporting observations were that only a certain maximum percentage of RBCs obtained from PNH patients could be hemolyzed by in vitro PNH hemolytic tests.

This suggested two populations of circulating RBCs in the PNH patient—one "resistant" (or normal) and one "sensitive" to hemolysis (i.e. PNH cells). In addition, selective sensitivity to hemolysis was demonstrated, both in vitro and in vivo, for a significant proportion of reticulocytes, suggesting that there are sensitive and resistant populations emerging from the marrow (2–5).

Strong evidence for the clonal theory of PNH stemmed from the work of Oni et al (14). In a black female PNH patient heterozygous for G-6-PD (i.e. type AB), the RBCs lysed in the acid hemolysis test were type B. It seems unlikely that the possession of G-6-PD type B per se renders the RBC more sensitive to the development of PNH, since (*a*) PNH has been commonly reported in Caucasians with a natural low incidence of type B, and (*b*) G-6-PD type A^- RBCs have been shown to be sensitive to PNH hemolysis (14, 15). In a hemizygous male PNH patient this followed a priori, and in a heterozygous female (type AA^-) this was demonstrated experimentally. In the hemizygous male there was no correlation between RBC G-6-PD activity and PNH hemolysis, i.e. as in uncomplicated PNH, the younger (vs older) RBCs were susceptible to hemolysis despite relatively more G-6-PD activity (15). Unfortunately for the sake of completeness, the residual unlysed RBCs in the patient of Oni et al (14) were not tested for G-6-PD type, and stronger in vitro hemolytic tests that might have lysed more RBCs were not employed (see Table 1).

It is clear that the low AChE of PNH RBCs is not a direct cause of hemolysis since accidental, suicidal, and deliberate experimental exposure to anticholinesterases has produced no evidence of hemolysis or of PNH erythrocytic defects (16). Nonetheless, using AChE deficiency as a possible clonal marker, Kunstling & Rosse showed that in PNH blood, aliquots of cells that hemolyzed in vitro had no AChE activity, whereas aliquots of nonhemolyzed cells had varying amounts of enzyme activity (17).

We attempted to define this on a unit-cellular basis, since aliquots of RBCs might obscure the presence of small populations. This was carried out by doubly staining the RBC for hemoglobin (to assess hemolysis) and for AChE following PNH hemolytic reactions. In Table 1, column 4 shows that $\leq 1.2\%$ of AChE-positive cells were hemolyzed. In PNH case 5, with 99% enzyme-negative cells in the circulating blood, at the most only one half hemolyzed in the acid hemolysis test. By using progressively stronger hemolytic tests (5% and 10% serum in sucrose solution), virtually all of the AChE-negative erythrocytes were hemolyzed. Similar findings were noted in PNH case 2, with 68% AChE-negative cells in circulating blood. Again, at the most only one half the AChE-negative erythrocytes were hemolyzed in the acid hemolysis test. Stronger hemolytic tests again reduced this number sharply toward zero. Figure 1 illustrates the staining characteristics of the RBC doubly stained for hemoglobin and for AChE. Preliminary marrow studies showed roughly the same percentage of AChE-deficient normoblasts as AChE-deficient circulating RBCs in given PNH patients.

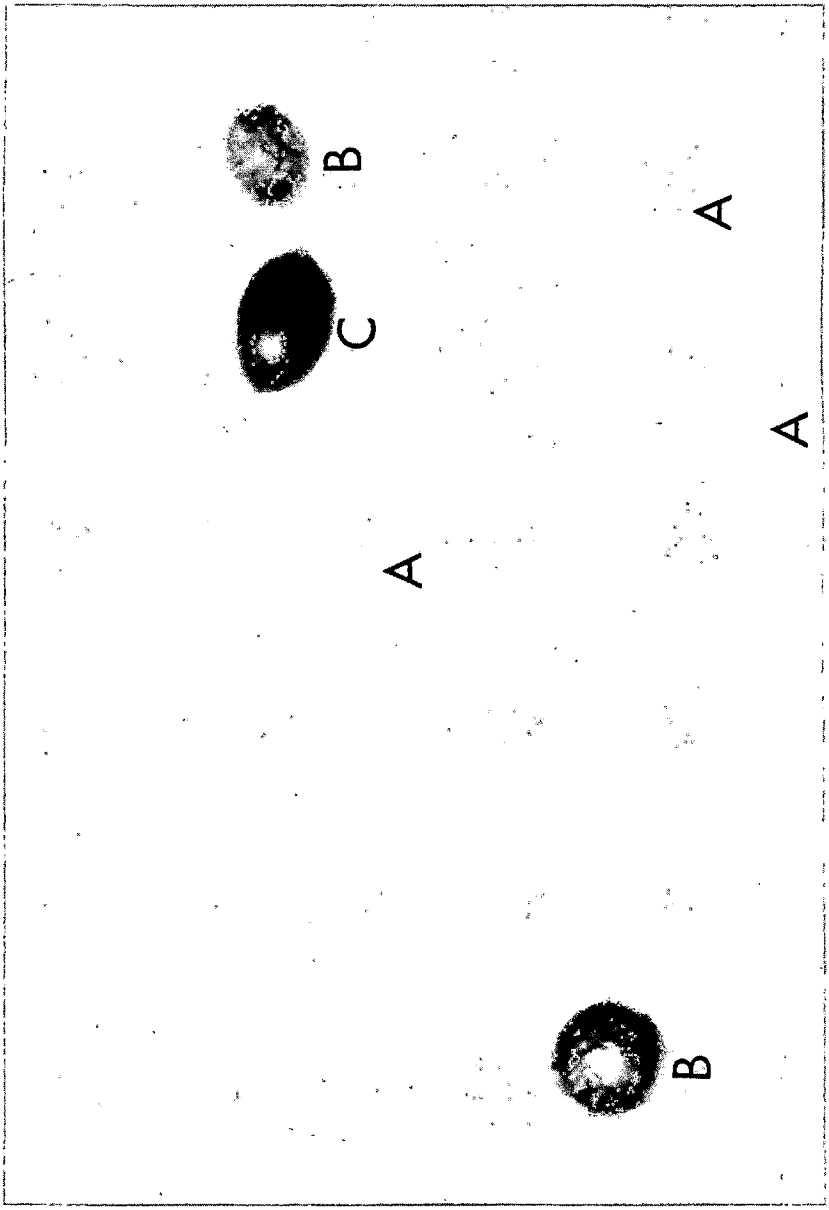
The clonal hypothesis and association of PNH with aplastic anemia, myeloid metaplasia, and leukemia has spurred interest in PNH hemolytic tests in these and other disorders with the aim of uncovering "hidden PNH clones." The association between PNH and aplastic anemia has been reviewed in detail (18). Acid hemolysis and, more particularly, sucrose hemolysis tests have been positive in associated disorders, although frequently weak and without other evidence of classical, full-blown PNH (19–21). The resulting enigma has been characterized by Lewis & Sirchia as "PNH: Disease or defect?" (22). In some instances there may actually be marrow clones producing PNH or PNH-like RBCs in these other disorders. In other instances, insignificant trace hemolysis and lack of proper controls may have led to technical overreading, falsely designated a "PNH clone."

Table 1 Correlation of erythrocyte AChE and hemolysis in individual PNH red cells[a]

	Before hemolysis	After hemolysis			
	AChE-negative	AChE-negative cells		AChE-positive cells	
	cells (%)	"Ghosts" (%)	Intact cells (%)	"Ghosts" (%)	Intact cells (%)
PNH Case 5					
Acid hemolysis	99.6	6.4	93.3	0	0.3
	98.8	50.8	48.8	0	0.4
Sucrose hemolysis:					
5% serum	99.1	96.8	2.4	0	0.8
	99.3	94.9	4.2	0.2	0.7
10% serum	99.4	98.4	0.8	0	0.8
	99.1	97.7	1.3	0	1.0
PNH Case 2					
Acid hemolysis	69.0	6.3	65.3	0.1	28.3
	72.4	29.2	28.1	0.4	42.3
Sucrose hemolysis:					
5% serum	70.3	57.0	9.3	0.4	33.3
	63.0	50.3	9.6	1.2	38.9
10% serum	64.9	59.7	4.5	1.2	34.6
	71.3	55.0	3.9	0.4	40.7

[a] Red cells from two PNH patients were stained for AChE activity before and after being subjected to the hemolytic tests designated. Hemolysis was determined by counterstaining the cells for hemoglobin. The procedures were carried out on each patient on two occasions one month apart. Each figure represents the average of duplicates of 1000 cell counts by two observers.

Figure 1 RBCs from reaction mixture following sucrose hemolysis test employing 5% serum in PNH Case 2. The RBCs were doubly stained for AChE and for hemoglobin. RBC AChE, a membrane-bound enzyme, is characterized by crystals or "needles" of reaction product. A= hemolyzed RBCs ("ghosts") without stainable AChE; B = intact RBCs with stainable AChE; and C= intact RBCs without stainable AChE. In separate experiments it was shown that >99% of RBCs from normal blood were AChE-positive and that AChE staining persisted after distilled-water hemolysis.


B
C
A
A
A
B

A third alternative deals with a possible degree of nonspecificity of the sucrose hemolysis test for PNH. Five to ten percent serum in sucrose solution brings about maximum hemolysis of PNH cells (23). Increasing the serum concentration in the reaction mixture negates low ionic strength, the key feature of the sucrose hemolysis test. Thus, the percentage of hemolysis falls as more than 10% serum is added to the system. If, however, ions are removed from the serum by preliminary dialysis against sucrose, the serum concentration in the reaction mixture can be increased above 10% without abolishing low ionic strength. In this system, increasing the concentration of serum in the sucrose solution above 10% not only augmented the percentage of PNH cells hemolyzed, but actually brought about striking hemolysis of normal cells (23). Thus, since the central feature concerned with lysis of PNH RBCs is exquisite sensitivity to complement hemolysis, and normal cells can be hemolyzed by manipulation of the sucrose hemolysis system (23), it is scarcely surprising that a variety of membrane defects might be associated with some increased sensitivity to complement and hence lead to positive PNH hemolytic tests on occasion.

Rosse and his collaborators have studied PNH cells with respect to complement sensitivity with a test employing anti-I antibody (24–26). They have described three populations of circulating RBCs occurring in various combinations in PNH patients: (*a*) markedly complement-sensitive cells, (*b*) slightly complement-sensitive cells, and (*c*) a complement-insensitive population indistinguishable from normal cells. However, using what may be more sensitive testing for PNH hemolysis and using RBC-AChE deficiency as a marker, we were able to obtain reasonably clear-cut separation into just two populations (Table I).

There have been relatively few cytogenetic studies in PNH, and many purport to show no abnormalities (4, 27). However, most have been with lymphocytes cultured from peripheral blood. Marrow karotype studies would appear more cogent. In six of nine patients, marrow studies showed variable but significant cytogenetic abnormalities (27). However, none demonstrated a Ph^1 chromosome as in CML. Likewise, there has been no study of chromosomal patterns in those few cases developing leukemia. One group has reported loss of the Y chromosome in an elderly PNH patient (27). Whether this denoted a pre-leukemic state or simply the effect of aging is not certain. The fact that not all cases of PNH showed marrow chromosomal changes in no way vitiates the concept of PNH as a clonal disorder, since chromosome abnormalities may occur only late in the course of clonal selection (7). More systematic and serially applied PNH marrow cytogenetic studies seem in order.

Observations that PNH leucocytes and platelets are also abnormal suggest a common stem cell disorder (4). However, unlike the situation with the PNH RBCs, there is no evidence of dual populations of PNH leucocytes and PNH platelets (4). This may, however, be due to the lack of adequate techniques. Unfortunately, this problem cannot be approached readily using AChE as a marker. Human platelets do not have cytochemically demonstrable AChE, and biochemical assessment yields only very low values. To date we have been unable to stain human leucocytes for AChE. Curiously, in PNH, as in CML, there is depressed leucocyte alkaline-phosphatase activity, but two populations of white cells have not been defined on this basis.

The clonal nature of PNH may provide a useful model for further study. As with CML, such studies may shed further light on the microevolutionary theory of "tumor progression in slow motion." PNH is a chronic disorder with an increasing number of patients with a greater than 10-year survival and with some "burned out" cases (2–5). Thus progression and even regression may provide clues to the study of clonal evolution. Obviously crucial to all of this will be the extension of studies to leucocytes as well as to marrow normoblasts in terms of AChE deficiency and complement sensitivity. Further study of PNH as a postulated clonal disorder may yield a fund of knowledge proportionately rich beyond the importance of this rare disorder in itself.

CONCLUSION

1. Clonal theories of disease, particularly progressive clonal growth and selection in tumorogenesis, were briefly cited.
2. Evidence for the clonal nature of PNH was presented. Correlation of red cell hemolysis with (*a*) G-6-PD type in two female G-6-PD mosaics with PNH and with (*b*) erythrocyte acetylcholinesterase deficiency, provides strong evidence for the clonal theory of PNH.
3. Possible pitfalls in defining "hidden PNH clones" in other diseases by the use of PNH hemolytic tests were discussed.
4. The potential of PNH as a study model for clonal evolution in human disease was emphasized.

ACKNOWLEDGMENT

Supported in part by Research Grant HL 18012 from the National Institutes of Health, Bethesda, Maryland, and by the Tampa Veterans Administration Hospital.

The authors wish to express appreciation to Dr. William M. LeFor, Chief of Transplant Laboratory, Veterans Administration Hospital and Assistant Professor of Medical Microbiology, University of South Florida, College of Medicine, and to Dr. Thomas A. Tedesco, Assistant Professor of Pediatrics and Director of Genetics Laboratory, University of South Florida College of Medicine for review of the manuscript and many helpful suggestions.

Literature Cited

1. Gartler, S. M. 1974. In *Chromosomes and Cancer,* ed. J. German, pp. 313–34. New York: Wiley. 756 pp.
2. Jensen, K. G., Killmann, S., eds. 1972. *Ser. Haematol.,* Vol. V, No. 3. Copenhagen: Munksgaard. 175 pp.
3. Rosse, W. F. 1972. In *Hematology,* ed. W. J. Williams, E. Beutler, A. J. Erslev, R. W. Rundles. pp. 460–74. New York: McGraw-Hill. 1480 pp.
4. Sirchia, G., Lewis, S. M. 1975. Paroxysmal nocturnal haemoglobinuria. *Clin. Haematol.* 4(1): 199–229
5. Dacie, J. V. 1967. *The Haemolytic Anaemias, Congenital and Acquired,* Pt. 4, pp. 1128–1260; pp. 993–1352. New York: Grune & Stratton. 2nd ed.
6. Nowell, P. C. 1974. See Ref. 1, pp. 267–85
7. Muldal, S., Lajtha, L. G. 1974. See Ref. 1, pp. 451–80
8. Adamson, J. W., Fialkow, P. J., Murphy, S., Prchal, J. F., Steinmann, L. 1975. Polycythemia vera: The nature of the cellular origin of the disease. *Proc. Ann. Meet. Am. Soc. Hematol., 18th, Dallas*
9. Lee, G. R., MacDiarmid, W. D., Cartwright, G. E., Wintrobe, M. M. 1968. Hereditary, X-linked, sideroachrestic anemia. The isolation of two erythrocyte populations differing in Xg^a blood type and porphyrin content. *Blood* 32:59–70
10. de Grouchy, J., Turleau, C. 1974. See Ref. 1, pp. 287–311
11. Dameshek, W. 1969. Forward and a proposal for considering paroxysmal nocturnal hemoglobinuria (PNH) as a "candidate" myeloproliferative disorder. *Blood* 33(2):263–64
12. Auditore, J. V., Hartmann, R. C., Flexner, J. M., Balchum, O. J. 1960. The erythrocyte acetylcholinesterase enzyme in paroxysmal nocturnal hemoglobinuria. *AMA Arch. Pathol.* 69: 534–43
13. Dacie, J. V. 1963. Paroxysmal nocturnal haemoglobinuria. *Proc. R. Soc. Med.* 56:587–96
14. Oni, S. B., Osunkoya, B. O., Luzzatto, L. 1970. Paroxysmal nocturnal hemoglobinuria: Evidence for monoclonal origin of abnormal red cells. *Blood* 36(2):145–52
15. Hartmann, R. C., Jenkins, D. E. Jr., Nance, W. E. 1965. PNH in a Negro male with G-6-PD deficiency. *Clin. Res.* 13:66 (Abstr.)
16. Hartmann, R. C., Jenkins, D. E. Jr. 1965. Paroxysmal nocturnal hemoglobinuria: Current concepts of certain pathophysiologic features. *Blood* 25(5): 850–65
17. Kunstling, T. R., Rosse, W. F. 1969. Erythrocyte acetylcholinesterase deficiency in paroxysmal nocturnal hemoglobinuria (PNH)—A comparison of the complement-sensitive and insensitive populations. *Blood* 33(4):607–16
18. Lewis, S. M., Dacie, J. V. 1967. The aplastic anaemia-paroxysmal nocturnal haemoglobinuria syndrome. *Br. J. Haematol.* 13:236–51
19. Hansen, N. E., Killmann, S.-Aa. 1970. Paroxysmal nocturnal hemoglobinuria in myelofibrosis. Blood 36(4):428–31
20. Kuo, C., Van Voolen, G. A., Morrison, A. N. 1972. Primary and secondary myelofibrosis: Its relationship to "PNH-like defect." *Blood* 40(6):875–80
21. Kruatrachue, M., Wasi, P., Na-Nakorn, S. 1974. Pattern of "paroxysmal nocturnal hemoglobinuria" red blood cell in aplastic anemia. *J. Med. Assoc. Thailand* 57(4):179–85
22. Lewis, S. M., Sirchia, G. 1972. PNH: Disease or defect? *Br. J. Haematol.* 23(Suppl.):71–81
23. Hartmann, R. C., Jenkins, D. E. Jr., Arnold, A. B. 1970. Diagnostic specificity of sucrose hemolysis test for paroxysmal nocturnal hemoglobinuria. *Blood* 35(4):462–75
24. Rosse, W. F., Dacie, J. V. 1966. Immune lysis of normal human and paroxysmal nocturnal hemoglobinuria (PNH) red blood cells. I. The sensitivity of PNH red cells to lysis by complement and specific antibody. *J. Clin. Invest.* 45(5):736–48
25. Rosse, W. F. 1973. Variations in the red cells in paroxysmal nocturnal haemoglobinuria. *Br. J. Haematol.* 24:327–42
26. Rosse, W. F., Adams, J. P., Thorpe, A. M. 1974. Populations of cells in paroxysmal nocturnal haemoglobinuria of intermediate sensitivity to complement lysis: Significance and mechanism of increased immune lysis. *Br. J. Haematol.* 28:181–90
27. Whang-Peng, J., Knutsen, T., Lee, E. C., Leventhal, B. 1976. Acquired XO/XY clones in bone marrow of a patient with paroxysmal nocturnal hemoglobinuria (PNH). *Blood* 47(4): 611–19

Ann. Rev. Med. 1977. 28:195–209
Copyright © 1977 by Annual Reviews Inc. All rights reserved

CLASSIFICATION OF THE HYPERLIPIDEMIAS[1]

❖7237

Richard J. Havel, M.D.
Cardiovascular Research Institute and Department of Medicine, University of California, San Francisco, California 94143

INTRODUCTION[2]

The subject of this review has received much attention during the last ten years. It has also been the object of considerable confusion and controversy. Much of the controversy is related to differing premises upon which classification has been based, and some to technical problems of lipoprotein analysis. Recent advances make it possible in some cases to replace plasma-phenotyping in individuals and kindred, with determination of abnormalities linked directly to abnormal gene products. In

---

[1] The author's research cited was supported by grants from the National Heart, Lung, and Blood Institute (HL-14237 and HL-06285).

[2] Glossary and abbreviations:

1. Abbreviations for lipoprotein classes and their electrophoretic equivalents are as follows: chylomicrons (same); very low-density lipoproteins—VLDL (pre-beta); low-density lipoproteins—LDL (beta); high-density lipoproteins—HDL (alpha).
2. The terms *hyperlipidemia* and *hyperlipoproteinemia* can be taken as synonymous, because essentially all the major plasma lipids are in macromolecular aggregates containing specific proteins. *Hyperlipemia* refers to conditions characterized by lactescence in serum or plasma obtained in the post-absorptive state owing to increased concentration of triglyceride-rich lipoproteins (chylomicrons and VLDL).
3. *Exogenous, endogenous,* and *mixed hyperlipemia* are generic terms denoting increased concentrations of chylomicrons (carrying dietary lipids), VLDL (mainly carrying hepatogenous lipids), and both, respectively.
4. *Hypercholesterolemia, hypertriglyceridemia,* and *combined hyperlipidemia* refer to elevations involving the major cholesterol-rich lipoproteins (LDL), triglyceride-rich lipoproteins (VLDL), and both, respectively. *Familial combined hyperlipidemia* also refers to a disorder in which LDL, VLDL, or both may be elevated—a better designation is probably *familial multiple lipoprotein-type hyperlipidemia.*
5. The other familial disorders discussed are *familial lipoprotein lipase deficiency, familial hypercholesterolemia* (owing to LDL-receptor abnormalities), and *familial dysbetalipoproteinemia* (denoting accumulation of remnant-like, triglyceride-rich lipoproteins).

the following, I summarize the developments that have led to our current state of knowledge, emphasizing the primary hyperlipidemias, and attempt to indicate the extent to which classification can assist in the management of patients with these disorders.

CLASSIFICATIONS BASED PRIMARILY UPON ANALYSIS OF PLASMA LIPIDS AND LIPOPROTEINS

The modern era of classification began with the publication by Gofman and his associates of a method to categorize individuals on the basis of absolute and relative concentrations of plasma lipoproteins of density <1.063 g/ml as determined by analytical ultracentrifugation (1). Subsequently, methods were developed for quantitative analysis of lipid and protein components of major lipoprotein classes separated by preparative ultracentrifugation (2, 3), thus providing for direct comparisons between the concentration of components such as cholesterol and triglycerides in individual lipoprotein classes and their concentration in serum (4, 5). With the development of direct methods for the measurement of triglycerides in serum (6,7), it became possible to relate easily the concentrations of serum cholesterol and triglycerides to the content of the major cholesterol-rich lipoproteins [low-density lipoproteins (LDL)—density limits between 1.006 and 1.063 g/ml] and the major triglyceride-rich lipoproteins [very low-density lipoproteins (VLDL)—density less than 1.006 g/ml]. These developments led to the first practical method for evaluating lipoprotein concentrations by relatively simple techniques to permit categorization of patients who had elevated concentrations of LDL (high cholesterol, normal triglycerides), of VLDL (high triglycerides, normal cholesterol), or combined elevation of these two lipoprotein classes (high cholesterol and triglycerides) that were considered to be associated with the development of atherosclerotic diseases (8). Particles representing dietary fat in transport (chylomicrons) could be detected by the formation of a creamy layer at the top of a stored sample of serum, particularly when it was stored in a refrigerator (3), whereas the other triglyceride-rich lipoprotein class (VLDL) remained evenly dispersed.

At about the same time as these developments occurred, it became evident that the major lipoprotein classes can also be separated by electrophoresis and that when stained with fat-soluble dyes after separation on a supporting medium such as paper, LDL have beta mobility and VLDL, pre-beta mobility (9, 10). The other major lipoprotein class, the high-density lipoproteins (HDL), has alpha-1 mobility; chylomicrons fail to migrate on paper because of their large size. Technical improvements in paper electrophoresis (11) led, in the mid-1960s, to a formal proposal for "phenotyping" familial hyperlipoproteinemias without use of the ultracentrifuge, based upon paper electrophoresis and measurement of cholesterol and triglyceride in blood serum obtained in the postabsorptive state (when chylomicrons are normally absent) (12). Shortly thereafter, the system was modified (13) and eventually it was expanded to include six types, three of which were characterized by elevation of a single lipoprotein class (types I, II-A, and IV), two by elevation of two classes (II-B and V), and one by an abnormal electrophoretic pattern with VLDL of beta rather than pre-beta mobility (14). This system became used widely and lipoprotein electrophoresis became a routine clinical tool along with quantitative analysis of serum cholesterol and triglycerides. Where hyperlipidemia was primary, this was accompanied by a widespread tendency to equate the phenotype determined in this manner with a specific genetic disorder. Although the originators of this typing system recognized that some of the phenotypes might be genetically heterogeneous, this tendency was reinforced by the proposal that specific diets are indicated in the treatment of each phenotype (15) and, perhaps, by the authority associated with Roman numerals as applied to genetic diseases, such as the glycogenoses. It rapidly became apparent, however, that a given phenotype seldom defines a single genotype. First, the phenotypes are mutable (16), a phenomenon to be expected particularly because of precursor-product relationships in the metabolism of VLDL and LDL and the large effects of diet on VLDL transport (16). Thus ingestion of large amounts of certain foods, particularly carbohydrates and ethanol, can increase VLDL levels substantially in ostensibly healthy, normolipidemic individuals (17–20), producing the phenotype "IV." Second, it was evident that at least one type (II) results not only from the dominantly transmitted disorder of familial hypercholesterolemia, but also from other, less well-defined disorders in which elevated levels of LDL are the primary expression (21). More recent investigations, to be described later, have greatly extended such observations, and it is now evident that in one of the most common familial groupings of hyperlipoproteinemia, elevations of LDL and VLDL, singly and in combination (types II-A, II-B, and IV) all occur with about equal frequency (22–24).

An additional problem with the original method used to define these six types is the inadequacy of paper electrophoresis. The patterns observed with the standard qualitative method are frequently nondiagnostic (25). One reason is that some individuals have a pre-beta band that usually migrates more slowly than VLDL and sediments at the density used to separate VLDL by flotation (26): "sinking pre-beta" (density limits 1.05–1.09 g/ml). This lipoprotein, which contains a distinct apoprotein [Lp (a)] and is evidently inherited as a quantitative trait (27), tends to be more prominent in hyperlipidemic subjects and in individuals with coronary heart disease (28, 29). In addition, recent studies have shown that many individuals have two pre-beta bands, both VLDL, one with the usual mobility and the other with slightly reduced mobility (30–32). The latter has characteristics suggesting that it may be a partially degraded triglyceride-rich lipoprotein ("remnant" particle), resembling the abnormal lipoprotein found in individuals with primary dysbetalipoproteinemia (type III), as described later (32). In such individuals the separation between beta and pre-beta lipoproteins may be indistinct, as it is in those with the type III ("broad beta") disorder, so that many samples labeled "type III" are not. A similar pattern may also be seen in hyperlipemic individuals at a time when triglyceride levels are falling rapidly.

Beyond these qualitative problems, paper electrophoresis and similar techniques using other supporting media such as agarose gel cannot, as ordinarily applied, be used to quantify lipoprotein concentrations. For all of these reasons, many research laboratories continue to use preparative ultracentrifugation to measure components (usually cholesterol and triglycerides, by automated methods) of the major lipoprotein classes (33–35).

Alternative approaches to quantitative evaluation of lipoprotein concentrations have been proposed. One is based upon estimation of the major lipoprotein classes separated by agarose-gel electrophoresis together with measurements of serum cholesterol and triglycerides (36, 37). Another uses membrane ultrafiltration and nephelometry, together with estimation of serum cholesterol, to assess LDL, VLDL, and chylomicrons in terms of small, medium, and large particle distributions with construction of an "SML profile" as the basis for phenotyping (38, 39). Generally speaking, the need for quantification of chylomicrons is limited and VLDL concentrations can almost always be estimated closely from the serum triglyceride level (35). A major purpose of these simpler systems, therefore, is estimation of the concentration of LDL ("beta quantification"). The precision of the simpler methods can usefully be compared with estimates that can be obtained from analysis of serum cholesterol and triglycerides alone. LDL-cholesterol, for example, can be estimated (33) as:

$$\text{serum cholesterol} - (\text{serum triglycerides}/5 + 45),$$

with a rank-order correlation coefficient of 0.89 (35). In this calculation, a normal composition of VLDL and an invariant value for HDL-cholesterol are assumed. The estimate of LDL-cholesterol from serum cholesterol and triglycerides can be improved (to a correlation coefficient on the order of 0.93) by using regression equations that take into account the systematic relationship between the concentration of HDL-cholesterol and the serum triglyceride concentration (35) (separate equations are used for the two sexes to take the higher HDL levels in women into account). Excellent precision can be achieved with this approach in subjects with serum triglycerides <400 mg/dl, by measuring HDL-cholesterol directly in serum after precipitation of all other major lipoprotein classes with heparin and Mn^{2+} (33, 40). When this is done, only a normal relationship between serum triglycerides and VLDL-cholesterol need be assumed. A related approach additionally utilizes sodium dodecyl sulfate to separate VLDL (41). Except for the last, none of these methods is satisfactory for those individuals with primary dysbetalipoproteinemia, because the cholesterol:triglyceride ratio in the abnormal VLDL of this disorder is closer to 1:1 than to the normal value of about 5:1.

An approach to classification of the various types of hyperlipidemia that ignores the problem of "beta-quantification" is based simply upon measurement of cholesterol and triglycerides in serum and detection of chylomicrons in refrigerated serum by their property of "creaming out" (42). It is particularly useful to indicate whether individuals have endogenous hyperlipemia (increased VLDL), exogenous hyperlipemia (increased chylomicrons), or mixed hyperlipemia (chylomicrons + VLDL) without the necessity of electrophoresis. With this method, the existence of accompanying increase in LDL is not specified, but can be estimated by one of the methods described above when hypercholesterolemia is also present.

In addition to the problems already enumerated, a major limitation of lipoprotein phenotyping is the arbitrary definition of normal limits. This is a necessitated by the fact that concentrations of LDL and VLDL are continuously distributed in human populations. Assignment of limits of normal that differ only slightly can greatly influence the relative as well as the absolute prevalence of each of the three major phenotypes that involve these two classes (43, 44). As described below, this fact complicates the analysis of inheritance of hyperlipidemias. In part, the problem is related to the large nongenetic influences upon lipoprotein concentrations and it has no apparent practical solution. For example, it is difficult to place individuals upon standard diets that maintain constant body weight for a substantial period before measurements are made. Even if this is done, the continuing influences of obesity still remain uncontrolled.

A quite distinct approach to phenotyping is based upon measurement of the concentration of specific apoproteins in plasma and in lipoproteins. Thus quantitative immunochemical analysis of apolipoprotein B (which is almost the only apoprotein in LDL as well as a major apoprotein component of VLDL) in plasma might better distinguish individuals with abnormalities that are variably expressed as elevations of VLDL, LDL, or both. For example, some individuals with endogenous hyperlipemia and normal levels of LDL have increased concentrations of apo-B in plasma, whereas others do not (45). However, this approach may not surmount the problem posed by nongenetic influences on lipid transport because they can influence apoprotein levels as well.

CLASSIFICATION BASED UPON THE FAMILIAL DISTRIBUTION OF LIPIDS AND LIPOPROTEIN CONCENTRATIONS

This approach to classification of primary hyperlipidemias antedates lipoprotein phenotyping and has much to commend it. In two disorders, the dominant form of familial hypercholesterolemia with tendon xanthomatosis and the Bürger-Grütz syndrome (42), identification of simple patterns of inheritance of discrete lipoprotein abnormalities has permitted a clear-cut nosology. These disorders, in which hyperlipidemia is invariably pronounced, account for only a small fraction of the primary hyperlipidemias. In most cases the hyperlipidemia is mild and variable in its expression. A major advance with this approach has been provisional definition of a new form of hyperlipidemia (22–24). In one extensive study of hyperlipidemic survivors of myocardial infarction (22), three disorders that were considered to be inherited as dominant traits were identified among about 60% of the kindred. Familial hypercholesterolemia, characterized by dominant inheritance of an approximately twofold increase in LDL level, and familial hypertriglyceridemia, in which increased levels of VLDL occurred in almost 50% of individuals older than 25 years, accounted for somewhat less than one half of these. In most of these families, hyperlipidemia occurred in almost 50% of members older than 25 years, but no single pattern predominated, with variable involvement of VLDL and LDL (familial combined hyperlipidemia). The inheritance pattern, taking all hyperlipidemic subjects together, was felt to be most consistent with a dominant trait, although reasons to question this conclusion have been presented (46). The syndrome of combined hyperlipidemia (also called multiple-type hyperlipoproteinemia (24)), was identified as a familial form of hyperlipidemia in two other studies reported about the same time (23, 24). In view of the metabolic interrelationship between VLDL and LDL mentioned earlier, it is not unreasonable for a single defect to be expressed differently, owing to variability of nongenetic influences upon rates of lipid transport or because of the influence of other genes upon the relationship.

Because of the problems posed by the continuous distribution of LDL and VLDL levels in the population and the important nongenetic influences upon lipoprotein concentrations, especially of VLDL, genetic analysis, except in the severe form of familial hypercholesterolemia with xanthomatosis, is difficult. In individual patients, secondary hyperlipidemias must of course be excluded. In doing so, it must be kept in mind that these may also be expressed variably. For example, hypothyroidism is usually accompanied by elevation of LDL, but triglyceride levels may also be increased, frequently with a predominance of particles that resemble remnant lipoproteins (47, 48). The accumulation of the remnant-like particles characteristic of primary dysbetalipoproteinemia is also very sensitive to thyroid status (49). Variable patterns are also seen in uncontrolled diabetes (50, 51) and in the nephrotic syndrome (4).

A more subtle problem is posed by the aggravation of hyperlipidemias produced by ingestion of ethanol or of various drugs. Both ethanol and contraceptive steroids, for example, increase VLDL levels moderately in ostensibly normolipidemic subjects, but they can cause striking hyperlipemia (endogenous or mixed) in subjects with preexisting mild endogenous hyperlipemia (52–55). Sever hyperlipemia can also occur when diabetes mellitus and familial endogenous hyperlipemia coexist (54). In these common situations, nature and nurture interact in a way that blurs the distinction between "primary" and "secondary" hyperlipidemia. It seems likely that similar interactions influence the extent to which a variety of secondary hyperlipidemias are expressed.

Not all primary forms of hyperlipidemia can be shown to be aggregated in families in the manner described above. Available data suggest that such "sporadic" or "polygenic" expressions of hyperlipidemias are more frequent than those that are "monogenic" (22), but systematic investigations of unselected populations in this respect have not been made.

CLASSIFICATION BASED UPON IDENTIFICATION OF ABNORMAL GENE PRODUCTS

For the familial disturbances that involve plasma lipoproteins primarily, the ultimate classification is of course one based upon identification of a gene product responsible for a specific abnormality. This aim appears to have been fulfilled in two cases thus far. First, the syndrome of familial exogenous hyperlipemia appears in almost all cases to result from deficient activity of the enzyme lipoprotein lipase, which catalyzes the rate-limiting step in plasma-triglyceride transport (56). The pattern of inheritance of this rare disorder is autosomal recessive (57). Second, the severe form of familial hypercholesterolemia appears to result from abnormalities of a high-affinity receptor on cell surfaces that recognizes certain classes of plasma lipoproteins, especially those containing the B-apoprotein, including LDL (58). Fibroblasts from heterozygous individuals generally have one half of the normally functional complement of these "LDL-receptors," but in some cases the receptor is defective, with a reduced capacity to bind LDL (59). The heterozygous condition is common, existing in 0.1–0.5% of the population (22, 57). The homozygous state is very rare, accompanied by much greater elevations of LDL and more severe and earlier atherosclerotic disease with extensive xanthomatosis (57). Some severe cases may represent combined heterozygotes for receptor-deficient and receptor-defective mutants (59).

A third disorder in which a single abnormal gene product now seems likely is primary familial dysbetalipoproteinemia (type III hyperlipoproteinemia, broad-beta or floating-beta disease). This disorder is probably more common than is generally realized, because its expression as hyperlipidemia generally requires the presence of obesity (60) or hypothyroidism (49). It has seldom been diagnosed in children and is rare in premenopausal females (57), perhaps because of a depressant effect of estrogens on the accumulation of the remnant-like VLDL particles that characterize this disorder (61). Various methods have been proposed to identify the abnormal VLDL. These include a ratio of cholesterol to triglycerides in VLDL exceeding 0.42 (62) and a ratio of VLDL-cholesterol to serum triglycerides exceeding 0.30 when serum triglyceride concentration is between 150 and 1000 mg/dl (63). Another approach is demonstration of the reduced (beta) mobility of the abnormal particle, but the distinction from individuals with an extra-slow pre-beta VLDL band is not always certain (32). Because all of these methods require preparative ultracentrifugation, others have been proposed (64–67), but they have not been widely used. The reduced mobility of the abnormal VLDL is the result of a lower content of certain apoproteins ("C" apoproteins) and increased content of an arginine-rich protein normally present in considerably smaller proportions in the protein moiety of VLDL (47, 68). This observation has raised the possibility that the content of this protein in VLDL or in serum might better distinguish the disorder, particularly if it were related to the concentration of other components of VLDL. Recently, it was reported that three electrophoretic components of the arginine-rich protein are separable by polyacrylamide-gel electrophoresis using the technique of isoelectric focusing (69) and that one of them is grossly deficient or absent in individuals with primary dysbetalipoproteinemia (70). This observation, which has been confirmed (Havel, unpublished observations), raises the possibility that this abnormality may reflect the abnormal gene product. If so, evaluation of close relatives of affected individuals should demonstrate whether, as has been proposed (60, 71), the disorder is transmitted as an autosomal dominant. At any rate, this apoprotein abnormality appears to provide an excellent new diagnostic tool and should help to define the prevalence of this not-uncommon disorder.

Identification of abnormal gene products would also determine whether the disorders defined as familial hypertriglyceridemia and familial combined hyperlipidemia are monogenic and dominantly transmitted or, alternatively, if they have a more complicated etiology. The fact that their expression is usually delayed until adult life suggests that abnormal regulation of a transport process may be involved. For example, VLDL levels in individuals with familial hypertriglyceridemia may be normal when rates of endogenous triglyceride transport are low. Either abnormalities of removal capacity or failure to adapt this capacity to increased transport rates may result in the expression of endogenous hyperlipemia. High-carbohydrate diets increase VLDL-triglyceride transport (72) and tend to produce a transient endogenous hyperlipemia in most individuals, which usually lasts for several weeks and then subsides (19), suggesting the possibility that an adaptive increase in removal capacity has occurred. Restriction of dietary carbohydrates, particularly if accompanied by weight loss in obese subjects with primary endogenous hyperlipemia, often results in a return to normal levels of VLDL, a result consistent with either of the possibilities mentioned above.

THERAPEUTIC CLASSIFICATION OF HYPERLIPIDEMIA

A major argument for separation of patients with primary hyperlipidemia according to the lipoprotein-phenotype is that it provides the basis for more effective therapeutic regimens based upon the lipoprotein class (or classes) involved. An approach to dietary treatment based upon phenotype has been devised from the substantial experience at the National Heart and Lung Institute (15, 73). Thus a lower intake of cholesterol is recommended for individuals with elevated levels of LDL than for those with increased VLDL. Different amounts of carbohydrates and simple sugars are also recommended for the various types, but except in the case of exogenous hyperlipemia, reduction of saturated fats and increased intake of polyunsaturated fats is advised for all. Other experienced therapists believe that a single "prudent" diet, restricted in cholesterol and saturated fats, with calories adjusted according to body size and the presence of obesity, provides a satisfactory approach to dietary management (74). A somewhat different approach that uses different diets for patients with elevated cholesterol than for those with elevated triglycerides (with or without elevated cholesterol) can be applied to all situations except the very rare familial exogenous hyperlipemia (16). Patients with elevated cholesterol are given a diet low in cholesterol and saturated fats. For those with elevated triglycerides, caloric restriction to achieve ideal weight on a balanced diet is used. Specific restriction of ethanol, and sometimes of simple sugars, may also be needed. This diet may be applied successfully to patients with four lipoprotein phenotypes, including those with mixed hyperlipidemia (chylomicrons plus VLDL). When cholesterol level remains elevated after weight reduction, cholesterol and saturated fats must of course be restricted. Unfortunately, no systematic comparison of these different dietary approaches to hyperlipemia has been made, but it is evident that the latter two do not require lipoprotein phenotyping.

Different drugs are indicated for patients with elevated cholesterol (LDL) or triglyceride (VLDL) levels that do not respond adequately to diet (16, 73, 75, 76). However, phenotyping does not usually help in the choice of drug and, in one situation, selection of drug solely on the basis of phenotype is inadequate. Thus, bile acid–binding resins are needed to reduce cholesterol levels to near normal in patients with heterozygous familial hypercholesterolemia (73, 76). This class of drugs is also effective in the many other patients with milder elevations of LDL, but in them *p*-chlorophenoxyisobutyrate (clofibrate) may also be quite effective (76) and is much easier to take, as well as cheaper. Clofibrate may also be used in diet-unresponsive patients with hypertriglyceridemia, but it may be ineffective in reducing concomitant hypercholesterolemia (75, 76) and an empirical approach must generally be used. Patients with primary dysbetalipoproteinemia (type III) are very responsive both to diet and clofibrate (73, 76), so that identification of this disorder is a guarantee of the efficacy of treatment. Most such patients that come to the attention of physicians have characteristic tuberoeruptive or planar xanthomata, so that the diagnosis can be established "bloodlessly" (16). If the diagnosis is missed by failure to apply the specialized techniques described above, the therapeutic approaches applicable to patients with combined hyperlipidemia will still be successful.

In one respect, classification that goes beyond simple measurement of serum cholesterol and triglycerides can be highly recommended. About 1% of men and 6% of women have HDL-cholesterol levels above 100 mg/dl [i.e. levels higher than are predicted from regression equations relating HDL-cholesterol to serum triglycerides (35)]. Such individuals obviously can have mild elevations of serum cholesterol with normal levels of LDL-cholesterol and serum triglycerides (individuals with triglyceride levels exceeding 200 mg/dl rarely have high HDL-cholesterol levels). Some evidence suggests that increased levels of HDL may be genetically determined (familial hyperalphalipoproteinemia) (77, 78) and evidence is emerging that high levels of HDL may be associated with reduced risk of atherosclerotic disease (79). If this is so, routine measurement of HDL-cholesterol, as well as serum cholesterol and triglycerides, may become routine in the "therapeutic classification" of hyperlipidemia. Measurement of HDL-cholesterol is simple and also permits reliable estimation of LDL-cholesterol in almost all individuals.

A final and fairly obvious point, with respect to the value of therapeutic classification of hyperlipidemia, is that it is the lipid or lipoprotein concentration that is decisive. Thus the decision to use drugs to treat hypercholesterolemia obviously depends upon the level of cholesterol in "atherogenic" lipoproteins (LDL and VLDL), recognizing that risk of atherosclerotic disease is a graded function of cholesterol level (80). The decision to use drugs to treat hypertriglyceridemia must take into account the conflicting evidence for the atherogenicity of isolated elevations of VLDL (80–83) and also the danger of pancreatitis when triglyceride levels (in VLDL and chylomicrons) exceed 1500–2000 mg/dl (42).

CURRENT STATUS OF CLASSIFICATION AND PROSPECTUS

For the physician responsible for identification and treatment of patients with hyperlipidemia, it now seems to be the consensus that relatively simple laboratory methods almost always suffice. Thus measurement of cholesterol and triglyceride in serum obtained in the postabsorptive state, with examination of refrigerated serum for chylomicrons, is generally considered to provide the basis for screening and for separation of patients into major therapeutic groups (84–88). Lipoprotein electrophoresis, formerly given great emphasis, is now felt to be unimportant (85, 86) or virtually irrelevant (87, 88) in clinical practice.

This change in emphasis reflects more than the technical limitations of lipoprotein electrophoresis. It recognizes that identification of specific lipoprotein classes is by itself an insufficient basis for classification (16, 43, 44, 84–91). Thus lipoprotein phenotyping has been replaced by emphasis on familial aggregations of hyperlipidemia, which have mainly a genetic basis.

However, identification of specific genetic disorders, except for familial lipoprotein lipase deficiency (with exogenous hyperlipemia), familial LDL-receptor abnormalities (with hypercholesterolemia), and possibly familial dysbetalipoproteinemia, remains difficult. Most patients with primary hypercholesterolemia or hypertriglyceridemia and those with combined hyperlipidemia cannot be assumed to have familial disorders in the absence of extensive investigations of close relatives (85, 88). Nor can the nature of a putative familial disorder be assumed when a patient with mild primary hypercholesterolemia or hypertriglyceridemia is identified. Thus, for most patients with primary hyperlipidemia, the proper designation would seem to be primary (hypercholesterolemia, hypertriglyceridemia, or combined hyperlipidemia), unclassified. Where possible, and particularly when severe hypercholesterolemia is present, systematic evaluation of lipid levels in close relatives is felt to be justified because prophylactic measures can then be applied (88).

It is widely recognized that evaluation of hyperlipidemia must take the various secondary forms into account, but due recognition must now be given to the fact that many patients who come to the attention of physicians, particularly with hypertriglyceridemia, have a primary hyperlipidemia that has been aggravated by additional factors (52–55, 92). Thus an appropriate classification may be primary hypertriglyceridemia, unclassified, associated with (obesity, alcoholism, diabetes mellitus, exogenous estrogen, etc).

From the therapeutic standpoint, the difficulty with respect to identification of a specific familial disorder is not necessarily unfortunate, because the current state of the art permits reasonably satisfactory management of the common "mild" hyperlipidemias on a generic basis (16, 74, 84, 92). This state of affairs may, however, change drastically in the future as the etiology of specific familial hyperlipidemias is uncovered. For example, the discovery of the LDL-receptor and its deficiency in the severe form of familial hypercholesterolemia has pointed to new possibilities for treatment that are more directly related to the genetic defect (58).

A recent review of the genetically determined hyperlipidemias provides a useful classification based upon present knowledge (92). Table 1 shows a pragmatic classification of primary hyperlipidemias that includes the frequent situation in which data from families are unavailable or do not permit more specific categorization. It must again be emphasized that "hypercholesterolemia" implies elevation of LDL and that measurement of HDL-cholesterol may be required to exclude hyperalphalipoproteinemia.

New discoveries of the etiology of hyperlipidemias are likely to include genetic disorders of the apoprotein components of lipoproteins as well as of their metabo-

Table 1 The primary hyperlipidemias

Type	Cholesterol level	Triglyceride level
Exogenous hyperlipemia		
familial lipoprotein lipase deficiency unclassified	normal to elevated	elevated (creamy layer, infranatant serum usually clear)
Endogenous and mixed hyperlipemia		
familial hypertriglyceridemia[a] familial multiple lipoprotein-type hyperlipidemia unclassified	normal to elevated	elevated (creamy layer with turbid infranatant serum defines mixed forms)
Hypercholesterolemia		
familial LDL-receptor deficiency familial LDL-receptor defect familial multiple lipoprotein-type hyperlipidemia polygenic hypercholesterolemia unclassified exogenous hypercholesterolemia[b]	elevated	normal
Combined hyperlipidemia		
familial multiple lipoprotein-type hyperlipidemia familial dysbetalipoproteinemia unclassified	elevated	elevated

[a] Some have defined separate endogenous and mixed forms of familial hypertriglyceridemia (57); diabetes mellitus usually accompanies the mixed form, which makes its distinct nature uncertain.
[b] Hypercholesterolemia owing to excessive dietary cholesterol.

lism. Additional receptors, particularly for remnant forms of triglyceride-rich lipoproteins, and the complex lipoprotein lipase system seem particularly attractive as loci of regulatory abnormalities. More research on the metabolism of lipoprotein classes that have not been directly linked to atherogenesis may show interactions that influence the concentration of VLDL and LDL. The extremely rare hyperlipidemia associated with deficiency of lecithin-cholesterol acyltransferase (93) may be a prototypical example. The association between a slow pre-beta lipoprotein [presumably the Lp (a) lipoprotein], coronary heart disease, and hyperlipidemia (28, 29, 94) could represent another. Further measurements of the concentration of these lipoproteins and of specific apoproteins in relation to familial aggregations of hyperlipidemia may provide new etiological clues. Finally, search for linkage of specific familial forms of hyperlipidemia to genetic markers may help to separate new genotypes.

Literature Cited

1. Gofman, J. W., Rubin, L., McGinley, J. P., Jones, H. B. 1954. Hyperlipoproteinemia. *Am. J. Med.* 17:514–20
2. Havel, R. J., Eder, H. A., Bragdon, J. H. 1955. The distribution and chemical composition of ultracentrifugally separated lipoproteins in human serum. *J. Clin. Invest.* 34:1345–53
3. Bragdon, J. H., Havel, R. J., Boyle, E. 1956. Human serum lipoproteins. I. Chemical composition for four fractions. *J. Lab. Clin. Med.* 48:36–42
4. Baxter, J. H., Goodman, H. C., Havel, R. J. 1960. Serum lipid and lipoprotein alterations in nephrosis. *J. Clin. Invest.* 39:455–65
5. Albrink, M. J. 1961. Lipoprotein pattern as a function of total triglyceride concentration of serum. *J. Clin. Invest.* 40:536–44
6. Van Handel, E., Zilversmit, D. 1957. Micromethod for the direct determination of serum triglycerides. *J. Lab. Clin. Med.* 50:152–57

7. Carlson, L. A., Wadström, L. B. 1959. Determination of glycerides in blood serum. *Clin. Chim. Acta* 4:197–205
8. Havel, R. J., Carlson, L. 1962. Serum lipoproteins, cholesterol and triglycerides in coronary heart disease. *Metabolism* 11:195–97
9. Dangerfield, W. G., Smith, E. B. 1955. An investigation of serum lipids and lipoproteins by paper electrophoresis. *J. Clin. Pathol.* 8:132–39
10. Smith, E. B. 1957. Lipoprotein patterns in myocardial infarction. Relationship between the components identified by paper electrophoresis and in the ultracentrifuge. *Lancet* 2:910–14
11. Lees, R. S., Hatch, F. T. 1963. Sharper separation of lipoprotein species by paper electrophoresis in albumin-containing buffer. *J. Lab. Clin. Med.* 61:518–28
12. Fredrickson, D. S., Lees, R. S. 1965. A system for phenotyping hyperlipoproteinemia. *Circulation* 31:321–27
13. Fredrickson, D. S., Levy, R. I., Lees, R. S. 1967. Fat transport in lipoproteins—An integrated approach to mechanisms and disorders. *N. Engl. J. Med.* 276:32–44, 94–103, 148–56, 215–25, 273–81
14. Beaumont, J. L., Carlson, L. A., Cooper, G. R., Fejfar, Z., Fredrickson, D. S., Strasser, T. 1970. Classification of hyperlipidaemias and hyperlipoproteinaemias. *Bull. WHO* 43:891–908
15. Fredrickson, D. S., Levy, R. I., Jones, E., Bonnell, M., Ernst, N. 1970. *The Dietary Management of Hyperlipoproteinemia: A Handbook for Physicians.* Washington: US Dept. Health, Educ. Welfare, Publ. Health Serv. 83 pp.
16. Havel, R. J. 1972. Hyperlipidemia: The significance and management. *Mod. Med.* 40:78–83
17. Watkin, D. M., Froeb, H. F., Hatch, F. T., Gutman, A. B. 1950. Effects of diet in essential hypertension. II. Results with unmodified Kempner rice diet in fifty hospitalized patients. *Am. J. Med.* 9:441–93
18. Hatch, F. T., Abell, L. L., Kendall, F. E. 1955. Effects of restriction of dietary fat and cholesterol upon serum lipids and lipoproteins in patients with hypertension. *Am. J. Med.* 19:48–60
19. Antonis, A., Bersohn, I. 1961. The influence of diet on serum-triglycerides in South African white and Bantu prisoners. *Lancet* 1:3–9
20. Jones, D. P., Losowsky, M. S., Davidson, C. S., Lieber, C. S. 1963. Effects of ethanol on plasma lipids in man. *J. Lab. Clin. Med.* 62:675–82
21. Patterson, D., Slack, J. 1972. Lipid abnormalities in male and female survivors of myocardial infarction and their first-degree relatives. *Lancet* 1:393–99
22. Goldstein, J. L., Schrott, H. G., Hazzard, W. R., Bierman, E. L., Motulsky, A. G. 1973. Hyperlipidemia in coronary heart disease. II. Genetic analysis of lipid levels in 176 families and delineation of a new inherited disorder, combined hyperlipidemia. *J. Clin. Invest.* 52:1544–68
23. Rose, H. G., Krantz, P., Weinstock, M., Juliano, J., Haft, J. I. 1973. Inheritance of combined hyperlipoproteinemia: Evidence for a new lipoprotein phenotype. *Am. J. Med.* 54:148–60
24. Nikillä, E. A., Aro, A. 1973. Family study of serum lipids and lipoproteins in coronary heart-disease. *Lancet* 1: 954–59
25. Pries, C., Van Gent, C. M., Baes, H., Polano, M. K., Hulsman, H. A. M., Querido, A. 1968. Primary hyperlipoproteinemia: The clinico-chemical classification of the most common types. *Clin. Chim. Acta* 19:181–90
26. Ellefson, R. D., Jimenez, B. J., Smith, R. C. 1971. Pre-β (or α_2) lipoprotein of high density in human blood. *Mayo Clin. Proc.* 46:328–32
27. Albers, J. J., Wahl, P., Hazzard, W. R. 1974. Quantitative gentic studies of the human plasma Lp (a) lipoprotein. *Biochem. Genet.* 11:475–86
28. Avogaro, P., Cazzolato, G. 1975. "Sinking" lipoprotein in normal, hyperlipoproteinaemic and atherosclerotic patients. *Clin. Chim. Acta* 61:239–46
29. Dahlén, G., Ericson, C., Furberg, C., Lundkvist, L., Svärdsudd, K. 1972. Studies on extra pre-beta lipoprotein fraction. *Acta Med. Scand.* Suppl. 531, pp. 1–29
30. Noble, R. P., Hatch, F. T., Mazrimas, J. A., Lindgren, F. T., Jensen, L. C., Adamson, G. L. 1969. Comparison of lipoprotein analyses by agarose gel and paper electrophoresis with analytical ultracentrifugation. *Lipids* 4:55–59
31. Carlson, L. A., Ericson, M. 1975. Quantitative and qualitative serum lipoprotein analysis. Part I. Studies in healthy men and women. *Atherosclerosis* 21:417–23
32. Havel, R. J., Kane, J. P., Pagnan, A. 1976. Metabolism of lipids and protein components of triglyceride-rich lipoproteins in rats and humans. In *Lipo-*

protein Metabolism, ed. H. Greten, pp. 65–68. Heidelberg: Springer. 160 pp.

33. Fredrickson, D. S. 1972. A physician's guide to hyperlipidemia. *Mod. Concepts Cardiovasc. Dis.* 41:31–36
34. National Heart and Lung Institute. 1974. Lipid Research Clinics Program *Manual of Laboratory Operations Vol. 1: Lipid and Lipoprotein Analysis.* Bethesda, Md: Publ. No. (NIH)75–628. 81 pp.
35. Myers, L. H., Phillips, N. R., Havel, R. J. 1976. Mathematical evaluation of methods for estimation of the concentration of the major lipid components of human serum lipoproteins. *J. Lab. Clin. Med.* 88:491–505
36. Hatch, F. T., Lindgren, F. T., Adamson, G. L., Jensen, L. C., Wong, A. W., Levy, R. I. 1973. Quantitative agarose gel electrophoresis of plasma lipoproteins: A simple technique and two methods for standardization. *J. Lab. Clin. Med.* 81:946–60
37. Lindgren, F. T., Adamson, G. L., Jensen, L. C., Wood, P. D. 1975. Lipids and lipoprotein measurements in a normal adult American population. *Lipids* 10:750–56
38. Stone, M. C., Thorp, J. M., Mills, G. L., Dick, T. B. S. 1970. Comparison of membrane filtration and nephelometry with analytical ultracentrifugation, for the quantitative analysis of low density lipoprotein fractions. *Clin. Chim. Acta* 30:809–28
39. Stone, M. C., Thorp, J. M., Mills, G. L., Dick, T. B. S. 1971. Diagnosis and classification of abnormal lipoprotein patterns. *Clin. Chim. Acta* 31:333–54
40. Friedewald, W. T., Levy, R. I., Fredrickson, D. S. 1972. Estimation of the concentration of low-density lipoprotein cholesterol in plasma, without use of the preparative ultracentrifuge. *Clin. Chem.* 18:499–502
41. Wilson, D. E., Spiger, M. J. 1973. A dual precipitation method for quantitative plasma lipoprotein measurement without ultracentrifugation. *J. Lab. Clin. Med.* 82:473–82
42. Havel, R. J. 1969. Pathogenesis, differentiation and management of hypertriglyceridemia. In *Advances in Internal Medicine* ed. G. H. Stollerman, 15:117–54. Chicago: Year Book Med. Publ. 494 pp.
43. Hazzard, W. R., Goldstein, J. L., Schrott, H. G., Motulsky, A. G., Bierman, E. L. 1973. Hyperlipidemia in coronary heart disease. III. Evaluation of lipoprotein phenotypes of 156 genetically defined survivors of myocardial infarction. *J. Clin. Invest.* 52:1569–77
44. Olsson, A. G., Carlson, L. A. 1975. Studies in asymptomatic primary hyperlipidaemia. I. Types of hyperlipoproteinaemias and serum lipoprotein concentrations, compositions and interrelations. *Acta Med. Scand.* Suppl. 580, pp. 1–37
45. Schonfeld, G., Lees, R. S., George, P. K., Pfleger, B. 1974. Assay of total plasma apolipoprotein B concentration in human subjects. *J. Clin. Invest.* 53:1458–67
46. Slack, J. 1973. A comparison of three recent reports on the inheritance of lipid abnormalities in survivors of myocardial infarction. In *Atherosclerosis III, Proc. Int. Symp. Atheroscl. 3rd,* ed. G. Schettler, A. Weizel, pp. 415–18. Berlin: Springer. 1034 pp.
47. Shore, B., Shore, V., Salel, A., Mason, D., Zelis, R. 1974. An apoprotein preferentially enriched in cholesteryl ester-rich very low density lipoproteins. *Biochem. Biophys. Res. Commun.* 58:1–7
48. Rössner, S., Rosenqvist, U. 1974. Serum lipoproteins and the intravenous fat tolerance test in hypothyroid patients before and during substitution therapy. *Atherosclerosis* 20:365–81
49. Hazzard, W. R., Bierman, E. L. 1972. Aggravation of broad-β disease (type 3 hyperlipoproteinemia) by hypothyroidism. *Arch. Intern. Med.* 130:822–28
50. Nikkila, E. A. 1973. Triglyceride metabolism in diabetes mellitus. In *Prog. Biochem. Pharmacol.: Effect of Carbohydrates on Lipid Metabolism*, ed. I. McDonald, 8:271–99. Basel: Karger. 310 pp.
51. Schonfeld, G., Birge, C., Miller, J. P., Kessler, G., Santiago, J. 1974. Apolipoprotein B levels and altered lipoprotein composition in diabetes. *Diabetes* 23:827–34
52. Mendelson, J. H., Mello, N. K. 1973. Alcohol-induced hyperlipidemia and beta lipoproteins. *Science* 180:1372–74
53. Ginsberg, H., Olefsky, J., Farquhar, J. W., Reaven, G. M. 1974. Moderate ethanol ingestion and plasma triglyceride levels: A study in normal and hypertriglyceridemic persons. *Ann. Intern. Med.* 80:143–49
54. Brunzell, J. D., Schrott, H. G. 1973. The interaction of familial and secondary causes of hypertriglyceridemia: Role in pancreatitis. *Trans. Assoc. Am. Physicians* 86:245–53

55. Pancreatitis from oral contraceptives. 1973. *Br. Med. J.* 4:688–89
56. Fielding, C. J., Havel, R. J. 1976. Lipoprotein lipase. *Arch. Pathol.* In press
57. Fredrickson, D. S., Levy, R. I. 1972. Familial hyperlipoproteinemia. In *The Metabolic Basis of Inherited Diseases,* ed. J. B. Stanbury, J. B. Wyngaarden, D. S. Frederickson, pp. 545–614. New York: McGraw-Hill. 1778 pp. 3rd ed.
58. Brown, M. S., Goldstein, J. L. 1976. Receptor-mediated control of cholesterol metabolism. Study of human mutants has disclosed how cells regulate a substance that is both vital and lethal. *Science* 191:150–54
59. Goldstein, J. L., Dana, S. E., Brunschede, G. Y., Brown, M. S. 1975. Genetic heterogeneity in familial hypercholesterolemia: Evidence for two different mutations affecting functions of low-density lipoprotein receptor. *Proc. Natl. Acad. Sci. USA* 72:1092–96
60. Morganroth, J., Levy, R. I., Fredrickson, D. S. 1975. The biochemical, clinical, and genetic features of type III hyperlipoproteinemia. *Ann. Intern. Med.* 82:158–74
61. Gagne, C., Kushwaha, R., Albers, J., Brunzell, J., Hazzard, W. 1975. Type III hyperlipidemia: Implications of paradoxical hypolipidemic response to estrogen. *Circulation* 52:Suppl. II, p. 39
62. Hazzard, W. R., Porte, D., Bierman, E. L. 1972. Abnormal lipid composition of very low density lipoproteins in diagnosis of broad-beta disease (type III hyperlipoproteinemia). *Metabolism* 21:1009–19
63. Fredrickson, D. S., Morganroth, J., Levy, R. I. 1975. Type III hyperlipoproteinemia: An analysis of two contemporary definitions. *Ann. Intern. Med.* 82:150–57
64. Seidel, D., Greten, H. 1970. Hyperlipoproteinämie—typ III: Eine möglichkeit zur immunologischen diagnostik. *Clin. Chim. Acta* 30:31–36
65. Masket, B. H., Levy, R. I., Fredrickson, D. S. 1973. The use of polyacrylamide gel electrophoresis in differentiating Type III hyperlipoproteinemia. *J. Lab. Clin. Med.* 81:794–802
66. Wieland, H., Seidel, D. 1973. Improved techniques for assessment of serum lipoprotein patterns. II. Rapid method for diagnosis of type III hyperlipoproteinemia without ultracentrifugation. *Clin. Chem.* 19:1139–41
67. Godolphin, W. J., Stinson, R. A. 1974. Isoelectrofocusing of human plasma lipoproteins in polyacrylamide gels: Diagnosis of type III hyperlipoproteinemia ("broad-beta" disease). *Clin. Chim. Acta* 56:97–103
68. Havel, R. J., Kane, J. P. 1973. Primary dysbetalipoproteinemia: Predominance of a specific apoprotein species in triglyceride-rich lipoproteins. *Proc. Natl. Acad. Sci. USA* 70:2015–19
69. Utermann, G. 1975. Isolation and partial characterization of an arginine-rich apolipoprotein from human plasma very-low-density lipoproteins: Apolipoprotein E. *Hoppe-Seylers Z. Physiol. Chem.* 356:1113–21
70. Utermann, G., Jaeschke, M., Menzel, J. 1975. Familial hyperlipoproteinemia type III: Deficiency of a specific apolipoprotein (Apo E-III) in the very-low-density lipoproteins. *FEBS Lett.* 56:352–55
71. Hazzard, W. R., O'Donnell, T. F., Lee, Y. L. 1975. Broad-β disease (type III hyperlipoproteinemia) in a large kindred. *Ann. Intern. Med.* 82:141–49
72. Havel, R. J. 1972. Caloric homeostasis and disorders of fuel transport. *N. Engl. J. Med.* 287:1186–92
73. Levy, R. I., Fredrickson, D. S., Shulman, R., Bilheimer, D. W., Breslow, J. L., Stone, N. J., Lux, S. E., Sloan, H. R., Krauss, R. M., Herbert, P. N. 1972. Dietary and drug treatment of primary hyperlipoproteinemia. *Ann. Intern. Med.* 77:267–94
74. Hall, Y., Stamler, J., Cohen, D. B., Mojonnier, L., Epstein, M. B., Berkson, D. M., Whipple, I. T., Catchings, S. 1972. Effectiveness of a low saturated fat, low cholesterol, weight-reducing diet for the control of hypertriglyceridemia. *Atherosclerosis* 16:389–403
75. Strisower, E. H., Adamson, G. L., Strisower, B. 1968. Treatment of hyperlipidemias. *Am. J. Med.* 45:488–501
76. Havel, R. J., Kane, J. P. 1973. Drugs and lipid metabolism. *Ann. Rev. Pharmacol.* 13:287–308
77. Glueck, C. J., Fallat, R. W., Millett, F., Gartside, P., Elston, R. C., Go, R. C. P. 1975. Familial hyper-alpha-lipoproteinemia: Studies in eighteen kindreds. *Metabolism* 24:1243–65
78. Avogaro, P., Cazzolato, G. 1975. Familial hyper-HDL-(α)-cholesterolemia. *Atherosclerosis* 22:63–77
79. Berg, K., Børresen, A-L., Dahlén, G. 1976. Serum-high-density-lipoprotein and atherosclerotic heart-disease. *Lancet* 1:499–501

80. Kannel, W. B., Garcia, M. J., McNamara, P. M., Pearson, G. 1971. Serum lipid precursors of coronary heart disease. *Hum. Pathol.* 2:129–51
81. Heyden, S. 1975. The problem with triglycerides. *Nutr. Metab.* 18:1–5
82. Carlson, L. A., Ekelund, L. G., Olsson, A. G. 1975. Frequency of ischaemic exercise E.C.G. changes in symptom-free men with various forms of primary hyperlipaemia. *Lancet* 2:1–3
83. Brunzell, J. D., Schrott, H. G., Motulsky, A. G., Bierman, E. L. 1976. Myocardial infarction in the familial forms of hypertriglyceridemia. *Metabolism* 25:313–20
84. Havel, R. J. 1975. Disorders of lipid metabolism. In *Textbook of Medicine*, ed. P. B. Beeson, W. McDermott, pp. 1629–37. Philadelphia: Saunders. 1892 pp. 14th ed.
85. Frederickson, D. S. 1975. It's time to be practical. *Circulation* 51:209–11
86. Iammarino, R. M. 1975. Lipoprotein electrophoresis should be discontinued as a routine procedure. *Clin. Chem.* 21:300–8
87. Havel, R. J. 1975. Hyperlipoproteinemia: Problems in diagnosis and challenges posed by the "type III" disorder. *Ann. Intern. Med.* 82:273–74
88. Motulsky, A. G., Boman, H. 1975. Screening for the hyperlipidemias. In *The Prevention of Genetic Disease and Mental Retardation*, ed. A. Milunsky, pp. 306–16. Philadelphia: Saunders. 506 pp.
89. Lewis, B. 1971. Classification of hyperlipidaemias. *Proc. R. Soc. Med.* 64: 905–8
90. Miettinen, T. A., Penttilä, I. M., Lampainen, E. 1972. Familial occurrence of mild hyperlipoproteinaemias. *Clin. Genet.* 3:271–80
91. Carlson, L. A. 1973. Clinical aspects of classification of primary hyperlipoproteinemias. In *Adv. Exp. Med. Biol.: Human Hyperlipoproteinemias—Principles and Methods*, ed. R. Fumagalli, G. Ricci, S. Gorini, 38:89–194. New York: Plenum. 292 pp.
92. Motulsky, A. G. 1976. Current concepts in genetics: The genetic hyperlipidemias. *N. Engl. J. Med.* 294: 823–27
93. Glomset, J. A., Norum, K. R. 1973. The metabolic role of lecithin:cholesterol acyltransferase: Perspectives from pathology. In *Adv. Lipid Res.*, ed. R. Paoletti, D. Kritchevsky, 11:1–65. New York: Academic. 363 pp.
94. Berg, K., Dahlén, G., Frick, M. H. 1974. Lp(a) lipoprotein and pre-β_1-lipoprotein in patients with coronary heart disease. *Clin. Genet.* 6:230–35

Ann. Rev. Med. 1977. 28:211–21
Copyright © 1977 by Annual Reviews Inc. All rights reserved

PULMONARY COMPLICATIONS OF GENERAL SURGERY

❖7238

Alan K. Pierce, M.D. and Joy Robertson, M.D.
Pulmonary Disease Division, Department of Medicine, University of Texas
Southwestern Medical School, Dallas, Texas 75235

Surgical patients who require general anesthesia may experience a variety of pulmonary complications. The seriousness of these problems is reflected in a report that 13.1% of postoperative deaths are secondary to pulmonary pathology (1). Since hypoxemia, the most frequent physiologic alteration due to lung disease in the perioperative period, exerts its effect on other organs, the morbidity and mortality secondary to pulmonary complications may be grossly underestimated.

The most frequent pulmonary complications of general surgery can be classified as hypoxemia, atelectasis, aspiration of gastric contents, thromboembolism, and respiratory failure in patients with preexisting severe pulmonary disease. Each of these complications is briefly reviewed.

HYPOXEMIA

The magnitude of hypoxemia following general surgery was first demonstrated by Nunn & Payne (2) in 1962. In a group of normal persons they found the mean arterial oxygen tension (Pa_{O_2}) while breathing room air to be 58 mm Hg in the immediate postoperative period; the lowest Pa_{O_2} recorded was 31 mm Hg. Significant hypoxemia persisted for the 24 hours of study, although no recognized pulmonary complications such as atelectasis existed. These results have been confirmed by other investigators (3, 4), and it has been estimated that hypoxia is responsible for half of the deaths associated with anesthesia (5).

Mechanisms of Hypoxemia

Although the data are not definitive (6, 7) it is likely that two interrelated mechanisms lead to hypoxemia in the perioperative period. One mechanism is diffuse microatelectasis, a process originally observed by Mead & Collier (8). These investigators noted that anesthetized dogs breathing in a normal tidal volume range had a progressive fall in lung compliance, which was rapidly reversed toward control values following forced inflations of the lung. The decreased tidal volume delivered by any given inflation pressure could be explained by multiple areas of atelectatic lung. The appearance of the lungs at postmortem examination supported this concept. They suggested that their findings could explain the decreased compliance previously reported in anesthetized compared to awake patients (9). Ferris & Pollard (10) demonstrated that a similar fall in compliance occurred in normal, unanesthetized humans who voluntarily breathed at normal but monotonously regular tidal volumes. Compliance returned to normal when the subjects were allowed to take large tidal volumes. Subsequently, microatelectasis has been related to the oxygenation of surgical patients (11–14). Patients who are not periodically hyperinflated to larger than normal resting tidal volumes become hypoxemic, but oxygenation, like compliance, can be restored to normal by large tidal volumes.

A second mechanism of hypoxemia during the perioperative period relates to differences in the regional distribution of ventilation in the lung. This process is closely related, but not precisely the same, as diffuse microatelectasis. The concept of regional differences in ventilation was originally proposed by West & Dollery (15) and has been studied in detail by investigators at McGill University (16–18). There is a gradient in pleural pressure down the lung so that the more superior parts of the lung are subjected to a more subatmospheric pressure. This gradient in pleural pressure is apparently partly gravity-dependent and is related to the weight and to the regional distortion of the lung. Thus, in an upright person, when the entire lung is at its resting volume [functional residual capacity (FRC)], apical lung units are at 60% of their total capacity, while units at the bottom of the lung are expanded to only 30% of their total capacity. At lung volumes less than FRC, the pleural pressure in dependent regions may exceed atmospheric pressure, leading to closure of airways in dependent lung regions with gas trapping behind the closed airways. The lung volume at which airways begin to close is referred to as the closing volume (CV) (19). Airway closure occurs very near residual volume in young, healthy persons, but occurs at larger lung volumes with advancing age (20).

The relationship between CV and FRC is an important determinant of matching of ventilation and blood flow, and hence an important determinant of arterial oxygenation. If airway closure occurs at a lung volume above FRC, the alveoli served by the closed airways do not receive normal ventilation. Functional residual capacity decreases from an upright to a supine position, due to an encroachment on the thorax by abdominal viscera (21); therefore this change is more marked in obese persons. The decrease in FRC is accentuated by surgical postures such as lithotomy (22). In addition, for reasons that are not clear, general anesthesia causes a reduction in FRC (23). This decrease in FRC occurs immediately after onset of anesthesia, is not progressive, remains postoperatively, and is not accompanied by a decrease in closing volume. Hence, dependent airway closure may contribute to arterial hypoxemia when an anesthetized or postoperative patient is ventilated with tidal volumes that are not large enough to reopen airways.

Premedication

Atropine is commonly administered preoperatively to minimize respiratory secretions. Although atropine causes several changes in pulmonary function (24), it does not lead to preoperative hypoxemia (24, 25). Similarly, preanesthetic doses of barbituates do not alter Pa_{O_2} (26). Preoperative narcotic agents, however, have been demonstrated to decrease Pa_{O_2} (26, 27) by suppressing spontaneous, periodic, large tidal volumes (27, 28). This finding suggests that preoperative narcotics should be avoided in hypoxemic patients or that oxygen should be administered from the time of preanesthesia. Additionally, all sedative drugs must be used with care in patients with chronic hypercapnia.

Anesthesia

Arterial hypoxemia is unlikely to occur during the course of anesthesia, since virtually all general anesthetic techniques include the administration of high concentrations of oxygen, which prevents hypoxemia despite the previously described changes in lung function. There is no apparent pulmonary advantage to any particular anesthetic technique (29–31). Halothane or ketamine may be associated with fewer episodes of acute attacks in asthmatic patients, but the evidence is not conclusive (32, 33).

Postoperative

Significant postoperative hypoxemia occurs predominantly in persons undergoing abdominal or thoracic surgery. There is a marked reduction in all static and dynamic lung volumes except for residual and closing volumes in persons undergoing laparotomies, but not in persons undergoing extremity surgery (34). Reductions in lung volumes are maximal on the first postoperative day and return to normal over approximately two weeks. These changes are ascribed to postoperative pain and a reduction in diaphragmatic motion (35). The resultant rapid, shallow breathing and unwillingness to breathe deeply or cough both lead to hypoxemia from airway closure and microatelectasis. Narcotic agents given to relieve postoperative pain do not prevent this sequence of events (27).

The high incidence of hypoxemia following abdominal and thoracic surgery suggests the need for oxygen administration to this particular group of postoperative patients. Although hypoxemia is not likely to be severe following non-abdominal surgery or procedures lasting less than 30 minutes (36), persons who might be adversely affected by hypoxemia (e.g. persons with arteriosclerotic heart disease) probably should be given oxygen at least for a while following such procedures. The means of oxygen administration is not critical in most patients. However, in patients with severe antecedent chronic obstructive lung disease who have the potential for hypoventilation, a precise low concentration of inspired oxygen should be administered by means of a mask that utilizes the Venturi principle.

ATELECTASIS

Atelectasis is the most commonly recognized pulmonary complication of general surgery. Fever and leukocytosis may be present due to growth of bacteria distal to an obstructed bronchus (37), but since symptoms and signs may regress with the removal of the bronchial obstruction without the use of antimicrobials, this syndrome usually is not considered a true pneumonia.

The reported incidence of postoperative atelectasis varies widely, depending on the diagnostic criteria; the most commonly used are fever, cough, increase in sputum production, and leukocytosis. By these standards the incidence has not changed significantly for many years, having been reported to be 8.9% in 1933 (38), 4.0% in 1949 (39), and 6.2% in 1968 (40). These statistics no doubt reflect the fact that older and more seriously ill patients are now considered surgical candidates.

Risk Factors

The type of operation performed markedly affects the likelihood of postoperative atelectasis. Operations that do not violate the peritoneum or pleura are associated with an incidence of less than 1% (39, 40). Such operations include not only head, neck, and extremity surgery, but also breast and anal surgery, as well as transurethral prostatic resections (40). Lower abdominal surgery is associated with approximately a 10% incidence of atelectasis, and upper abdominal procedures with a 20% incidence. The type of abdominal surgery performed is evidently not important.

Men have been thought to have a higher incidence of postoperative atelectasis than women, but the difference disappears among patients with similar smoking histories (40). Older persons (36, 40) and persons who are significantly overweight (41) are more likely to have postoperative complications, as are patients who are operated on in the presence of an upper respiratory infection (42).

It has already been indicated that postoperative narcotics do not allow patients to breathe deeply or cough (27). These changes have been shown to be dose-related. Bunker et al (43) found that men given more than four doses of narcotic following upper abdominal surgery were more likely to develop atelectasis than men given four or fewer doses.

An increased incidence of postoperative atelectasis among smokers has been noted for years. The amount of cigarette smoking is an important determinant. When patients were graded from nonsmokers to very heavy smokers, there was a stepwise increase in the incidence of complications from 8 to 43% (44). This correlation existed even though pulmonary functions, as indicated by the forced expiratory volume in one second (FEV_1) had not been impaired among the heavy smokers.

The most important risk factor is the presence of chronic obstructive lung disease (45–47). Stein et al (46) have clearly demonstrated the predictive value of pulmonary-function testing for postoperative pulmonary complications. These investigators studied patients preoperatively with a battery of pulmonary-function tests. Among patients with normal pulmonary functions, the incidence of atelectasis was 3%, but among patients with one or more abnormal tests, the incidence was 70%. The most sensitive test was the measurement of maximal expiratory flow rate, although changes in FEV_1 also tended to correlate well with morbidity.

Prevention and Treatment

Therapy in the postoperative period should include maximal inspiratory maneuvers combined with efforts to clear secretions from the tracheobronchial tree. Regimens incorporating these principles would best reverse the previously described physiologic changes that result in hypoxemia and atelectasis. Several modes of therapy have been suggested to achieve this goal. These include blow bottles, inhalation of carbon dioxide, intermittent positive-pressure breathing (IPPB), incentive spirometers, indwelling tracheal catheters, chest physiotherapy, and inhaled bronchodilators. Blow bottles or blowing out against some other expiratory resistance increases pleural pressure surrounding airways. The positive intrapleural pressure necessary to overcome the expiratory resistance at the mouth has a tendency to reduce the caliber of the airways, especially at low lung volumes, a result contrary to the desired physiologic effects. It is conceivable that if a patient inhales to a large lung volume before forced exhalation, a net beneficial effect occurs. However, blow bottles have never been investigated by a controlled clinical trial (48).

Stimulation of respiration by the inhalation of increased carbon dioxide concentrations has been attempted (49, 50). Since the production of large tidal volumes by this method is erratic, it may not result in reversal of the physiologic alterations found in the postoperative patient (51). Inhalation of carbon dioxide clearly should be avoided in patients with obstructive lung disease. Oxygen should be concurrently administered to all patients. These factors make carbon dioxide inhalation an unattractive procedure, in our opinion.

The routine use of IPPB for prevention of postoperative atelectasis was first suggested in 1958 (52). Subsequently, four controlled studies have evaluated postoperative IPPB (53–56). Three studies reported no benefit and one reduced the incidence of atelectasis by 17%. The latter study differed by employing preoperative instruction, higher inflation pressures, and larger volumes of bronchodilators (56). None of these studies recorded the most important variable determining good therapy, the tidal volume achieved during treatments.

Therapy with IPPB would be expected to be beneficial only if a large inspiratory volume is induced. It is reasonable to believe that such large tidal volumes are not obtained by routinely administered treatments. Therefore, routine use of IPPB likely has little benefit. If, however, the patient cannot be encouraged to inspire maximally spontaneously, the skilled administration of IPPB with monitoring of exhaled tidal volume could be visualized as a reasonable approach. Indirect evidence suggests that tidal volumes of at least 50% of the patient's predicted vital capacity must be administered to achieve a reversal of microatelectasis (57). It is likely that preoperative instruction is necessary to achieve this goal postoperatively.

Recently a device intended to encourage patients in taking maximal voluntary inspirations, an incentive spirometer, has been reported (58–60). The patient is rewarded with a light when a predetermined tidal volume is achieved. A decrease in postoperative complications was reported, using insufficiently tabulated data. If it is confirmed that this device assists deep breathing while minimizing personnel time, its use seems rational.

An entirely different approach to the treatment of postoperative atelectasis was reported in 1960 (61). The procedure consists of placement of a percutaneous plastic endotracheal catheter at the time of surgery with subsequent instillations of solutions to stimulate involuntary cough. As would be expected, varying results have been reported, from complete absence of postoperative complications to no statistical benefit (61–63). Data are insufficient to evaluate this procedure, but its use in asthmatic patients should be avoided, since tracheal stimulation may evoke bronchospasm in these patients.

The beneficial effects obtained with physiotherapeutic maneuvers have been well demonstrated by Thoren (64). Physiotherapy is defined as instruction of the patients in deep breathing, encouragement of cough, and postural drainage. Reduction of radiographic evidence of atelectasis from 42% to 12% was demonstrated in a series of patients undergoing laparotomies when treatment was given twice daily postoperatively after adequate preoperative instruction (64).

An additional advantage of inhaled bronchodilators has been demonstrated (65). Radiographic evidence of atelectasis was reduced from 43% in a group of patients receiving breathing training alone to 9% when bronchodilators and breathing training were combined. Likewise, high-risk patients with abnormal pulmonary-function tests also had a reduction of postoperative complications when inhaled bronchodilators and physiotherapy were combined (46, 66).

ASPIRATION OF GASTRIC CONTENTS

Regurgitation of gastric juice into the pharynx has been observed in up to one fourth of patients undergoing general anesthesia, and into the trachea in up to one sixth (67, 68). This complication results in one tenth of all anesthetic deaths (69), and this fraction is higher in obstetrical anesthesia.

The consequences of gastric contents aspiration depend on the volume and the pH of the aspirated material (70). If a large volume of material with a pH of 2 or less is aspirated, a diffuse, hemorrhagic pulmonary edema ensues. The term adult respiratory distress syndrome is frequently used to describe this noncardiogenic pulmonary edema.

The prevention of this potentially serious complication is primarily a matter of anesthesia technique. The pathogenesis and treatment of noncardiogenic pulmonary edema is too complex to be covered in this limited review; therefore we shall not deal with this subject further.

THROMBOEMBOLISM

It has long been recognized that postoperative patients have a high incidence of venous thrombosis and pulmonary embolism. Utilizing the ^{125}I-labeled fibrinogen technique, it has been demonstrated that up to 74% of postoperative patients have deep-venous thrombosis (71–73); autopsy results demonstrate deep-vein thrombosis in over 30% of patients dying after surgery (74), and pulmonary embolism is found in a similar percentage (75). Patients most at risk are those with trauma, especially fractures of the hip; older patients; those with preexisting venous insufficiency; and those with malignant neoplasms (76). Thromboembolism is a particularly dangerous complication because its occurrence is unpredictable and its manifestations may be so deceivingly nonspecific as to escape diagnosis before a fatal event (77). Accurate diagnosis depends upon a variety of laboratory and radiographic determinations, including pulmonary angiography, radioactive lung scanning, venography, ^{125}I-fibrinogen scanning, ultrasound techniques, and measurement of fibrin degradation products in serum.

The mainstay of therapy is anticoagulation; the most appropriate method is intravenously administered heparin, but oral therapy with warfarin may be useful as well, especially for long-term management (78, 79). Recently, interest has focused on prevention of postoperative thromboembolic complications utilizing a variety of techniques, including low-dose heparin (71); therapeutic doses of warfarin, dextran, and dipyridamole (Persantin) (80); pressure bandages or stockings on the lower extremities (81); and intermittent calf compression (76). Of these, low-dose or "mini-dose" heparin (5000 units subcutaneously every 8 or 12 hr) has been best shown to have a significant effect in reducing the incidence not only of ^{125}I-fibrinogen–detectable venous thrombosis, but of fatal pulmonary embolism as well (82).

RESPIRATORY FAILURE

Ventilatory Failure

Patients with severe chronic obstructive lung disease are at risk of developing postoperative ventilatory failure due to increased metabolic needs, necessitating a 25% increase in ventilation (83), narcotic administration, and the pulmonary changes indicated in a previous section that occur among these as well as normal persons. The studies by Miller, Wu & Johnson (84) and by Boutros & Weisel (30) suggest that ventilatory failure should be anticipated if preoperative spirometry indicates that maximal expiratory flow rates have been reduced to 15% of predicted values. Additionally, patients with preoperative hypercapnia have been demonstrated to develop postoperative ventilatory failure (46, 84).

Patients with severe pulmonary impairment are treated preoperatively by bronchodilators and physiotherapy until expectorated secretions are minimal. The preanesthetic regimen should not include narcotics. We prefer that the anesthetic regimen include orotracheal intubation so that postoperative ventilatory support, if needed, is facilitated. The endotracheal tube and ventilatory support are continued postoperatively until the patient is fully awake. An assessment of the patient's ventilatory capacity is made, and if it is apparently adequate, the patient is observed while breathing spontaneously. If the patient is alert, vital signs are stable, and blood-gas values do not deteriorate seriously, the patient may be extubated. If the patient deteriorates without ventilatory support, a regimen similar to that for any other patient in ventilatory failure is maintained. The best available data suggest that patients with severe pulmonary disability who are managed carefully in the perioperative period do not have a greater mortality than patients with no known lung disease receiving standard care (30, 85).

Pulmonary Hypertension

Severe pulmonary hypertension likely constitutes a strong contraindication to general surgery. This criterion is suggested by the 27% mortality rate of pregnant women with the Eisenmenger syndrome and 53% mortality rate of women with primary pulmonary hypertension at time of delivery (86). Further, Skinner & Pearce (87) found a 50% mortality for intrathoracic or intraabdominal surgery among patients with chronic obstructive lung disease complicated by cor pulmonale. Although definitive data are not available, it is likely that the increased mortality of these patients occurs during, rather than after surgery. Peripheral vasodilation during general anesthesia may cause hypotension due to the limited cardiac output imposed by the pulmonary hypertension.

CONCLUSIONS

1. Patients should be divided preoperatively into low- or high-risk categories, depending on their probability of developing postoperative pulmonary complications. The evaluation should include spirometry as well as an assessment of the previously defined risk factors.
2. Patients in a low-risk category need only instruction in deep breathing pre- and postoperatively. Routine use of supplemental oxygen postoperatively is reasonable until it can be demonstrated whether such is necessary.
3. High-risk patients should be as free as possible of respiratory secretions at the time of surgery. A regimen for this purpose includes cessation of smoking, and administration of inhaled bronchodilators followed by chest percussion and postural drainage.
4. High-risk patients should be carefully instructed in deep breathing and coughing preoperatively. A mechanical device such as an incentive spirometer may be beneficial in this regard. If it is not possible to achieve spontaneous deep breathing, an attempt to accomplish this by IPPB may be undertaken. The tidal volume desired should be ordered. If IPPB does not result in large tidal volumes, it should be discontinued.
5. The deep-breathing procedure found to be most successful preoperatively should be continued postoperatively.
6. The patient should be as mobile as possible while in bed and ambulated as soon as is feasible.
7. Patients with preoperative expiratory flows of less than 20% of predicted values or with chronic hypercapnia should be carefully observed for postoperative ventilatory failure.

Literature Cited

1. Vossschulte, K. 1958. Die postoperativen lungenkomplikationen ohne berucksichtigung der tuberdulose. *Arch. Klin. Chir.* 288:328–43
2. Nunn, J. F., Payne, J. P. 1962. Hypoxaemia after general anesthesia. *Lancet* 2:631–32
3. Diament, M. L., Palmer, K. N. V. 1967. Venous/arterial pulmonary shunting as the principal cause of postoperative hypoxemia. *Lancet* 1:15–17
4. Thompson, D. S., Eason, C. N. 1970. Hypoxemia immediately after operation. *Am. J. Surg.* 120:649–51
5. Bendixen, H. H., Laver, M. B. 1965. Hypoxia in anesthesia. *Clin. Pharmacol. Ther.* 6:510–39
6. Marshall, B. E., Wyche, M. Q. 1972. Hypoxemia during and after anesthesia. *Anesthesiology* 37:178–209
7. Rehder, K., Sessler, A. D., Marsh, H. M. 1975. General anesthesia and the lung. *Am. Rev. Respir. Dis.* 112:541–63
8. Mead, J., Collier, C. 1959. Relation of volume history of lungs to respiratory mechanics in anesthetized dogs. *J. Appl. Physiol.* 14:669–78
9. Nims, R. G., Conner, E. H., Comroe, J.

H. 1955. The compliance of the human thorax in anesthetized patients. *J. Clin. Invest.* 34:744–50

10. Ferris, B. G., Pollard, D. S. 1960. Effect of deep and quiet breathing on pulmonary compliance in man. *J. Clin. Invest.* 39:143–49
11. Egbert, L. D., Laver, M. B., Bendixen, H. H. 1963. Intermittent deep breaths and compliance during anesthesia in man. *Anesthesiology* 24:57–60
12. Bendixen, H. H., Hedley-Whyte, J., Chir, B., Laver, M. B. 1963. Impaired oxygenation in surgical patients during general anesthesia with controlled ventilation: a concept of atelectasis. *N. Engl. J. Med.* 269:911–96
13. Hedley-Whyte, J., Laver, M. B., Bendixen, H. H. 1964. Effect of changes in tidal ventilation on physiologic shunting. *Am. J. Physiol.* 206:891–97
14. Bendixen, H. H., Bullwinkel, B., Hedley-Whyte, J., Laver, M. B. 1964. Atelectasis and shunting during spontaneous ventilation in anesthetized patients. *Anesthesiology* 25:297–301
15. West, J. B., Dollery, C. T. 1960. Distribution of blood flow and ventilation-perfusion ratio in the lung, measured with radioactive CO_2. *J. Appl. Physiol.* 15:405–10
16. Ball, W. C., Stewart, P. B., Newsham, L. G. S., Bates, D. V. 1962. Regional pulmonary function studied with xenon^{133}. *J. Clin. Invest.* 41:519–31
17. Kaneko, K., Milic-Emili, J., Dolovich, M. B., Dawson, A., Bates, D. V. 1966. Regional distribution of ventilation and perfusion as a function of body position. *J. Appl. Physiol.* 21:767–77
18. Milic-Emili, J., Henderson, J. A. M., Dolovich, M. B., Trop, D., Kaneko, K. 1966. Regional distribution of inspired gas in the lung. *J. Appl. Physiol.* 21:749–59
19. McCarthy, D. S., Spencer, R., Greene, R., Milic-Emili, J. 1972. Measurement of "closing volume" as a simple and sensitive test for early detection of small airway disease. *Am. J. Med.* 52:747–53
20. Buist, A. S., Ross, B. B. 1973. Predicted values for closing volumes using a modified single breath nitrogen test. *Am. Rev. Respir. Dis.* 107:744–52
21. Don, H. F., Craig, D. B., Wahba, W. M., Couture, J. G. 1971. The measurement of gas trapped in the lungs at functional residual capacity and the effects of posture. *Anesthesiology* 35:582–90
22. Craig, D. B., Wahba, W. M., Don, H. 1971. Airway closure and lung volumes in surgical positions. *Can. Anaesth. Soc. J.* 18:92–99
23. Rehder, K., Hatch, D. J., Sessler, A. D., Marsh, H. M., Fowler, W. S. 1971. Effects of general anesthesia, muscle paralysis, and mechanical ventilation on pulmonary nitrogen clearance. *Anesthesiology* 35:591–601
24. Don, H. F., Robson, J. G. 1965. The mechanics of the respiratory system during anesthesia: the effects of atropine and carbon dioxide. *Anesthesiology* 26:168–78
25. Medrado, V., Stephen, C. R. 1966. Effect of premedication with atropine sulphate on arterial blood-gases and pH. *Lancet* 1:734–35
26. Pierce, J. A., Garofalo, M. L. 1965. Preoperative medication and its effect on blood gases. *J. Am. Med. Assoc.* 194:99–102
27. Egbert, L. D., Bendixen, H. H. 1964. Effect of morphine on breathing pattern, a possible factor in atelectasis. *J. Am. Med. Assoc.* 188:485–87
28. Bendixen, H. H., Smith, G. M., Mead, J. 1964. Pattern of ventilation in young adults. *J. Appl. Physiol.* 19:195–98
29. Boutros, A. R. 1965. Effects of nitrous oxide—curare, ether and cyclopropane on postoperative respiratory inadequacy. *Anesthesiology* 26:743–50
30. Boutros, A. R., Weisel, M. 1971. Comparison of effects of three anesthetic techniques on patients with severe pulmonary obstructive disease. *Can. Anaesth. Soc. J.* 18:286–92
31. Bridenbaugh, P. O., Bridenbaugh, L. D., Moore, D. C., Thompson, G. E. 1972. The role of intercostal block and three general anesthetic agents as predisposing factors to postoperative pulmonary problems. *Anesth. Analg.* 51:638–44
32. Gold, M. I. 1970. Anesthesia for the asthmatic patient. *Anesth. Analg.* 49:881–88
33. Corssen, G., Gutierrez, J., Reves, J. G., Huber, F. C. 1972. Ketamine in the anesthetic management of asthmatic patients. *Anesth. Analg.* 51:588–96
34. Anscombe, A. R., Buxton, R. St. J. 1958. Effect of abdominal operations on total lung capacity and its subdivisions. *Br. Med. J.* 2:85–86
35. Lewis, F. J., Welch, J. A. 1965. Respiratory mechanics in postoperative patients. *Surg. Gynecol. Obstet.* 120:305–14
36. Marshall, B. E., Millar, R. A. 1965.

Some factors influencing postoperative hypoxaemia. *Anesthesia* 20:408–28

37. Lansing, A. M., Jamieson, W. G. 1968. Mechanisms of fever in pulmonary atelectasis. *Arch. Surg.* 87:168–74

38. King, D. S. 1933. Postoperative pulmonary complications, I. A statistical study based on two years' personal observation. *Surg. Gynecol. Obstet.* 56: 43–50

39. Pooler, H. E. 1949. Relief of postoperative pain and its influence on vital capacity. *Br. Med. J.* 2:1200–3

40. Wightman, J. A. K. 1968. A prospective survey of the incidence of postoperative pulmonary complications. *Br. J. Surg.* 55:85–91

41. Cahill, J. M. 1968. Respiratory problems in surgical patients. *Am. J. Surg.* 116:362–68

42. Steward, D. J., Stoan, I. A. J. 1969. Recent upper respiratory infection and pulmonary artery clamping in the aetiology of postoperative respiratory complications. *Can. Anaesth. Soc. J.* 16:57–60

43. Bunker, J. P., Bendixen, H. H., Sykes, M. K., Todd, D. P., Surtees, A. D. 1959. A comparison of ether anesthesia with thiopental—nitrous oxide—succinylcholine for upper abdominal surgery. *Anesthesiology* 20:745–52

44. Chalon, J., Tayyab, M. A., Ramanathan, S. 1975. Cytology of respiratory epithelium as a predictor of respiratory complications after operation. *Chest* 67:32–35

45. Meneely, G. R., Ferguson, J. L. 1961. Pulmonary evaluation and risk in patient preparation for anesthesia and surgery. *J. Am. Med. Assoc.* 175:1074–80

46. Stein, M., Koota, G. M., Simon, M., Frank, H. A. 1962. Pulmonary evaluation of surgical patients. *J. Am. Med. Assoc:* 181:765–70

47. Lattimer, R. G., Dickman, M., Day, W. C., Gunn, M. L., Schmidt, C. D. 1971. Ventilatory patterns and pulmonary complications after upper abdominal surgery determined by preoperative and postoperative computerized spirometry and blood gas analysis. *Am. J. Surg.* 122:622–32

48. Bartlett, R. H., Gazzaniga, A. B., Geraghty, T. R. 1973. Respiratory maneuvers to prevent postoperative pulmonary complications: a critical review. *J. Am. Med. Assoc.* 224:1017–21

49. Schwartz, S. E., Dale, W. A., Rahn, H. 1957. Dead-space re-breathing tube for prevention of atelectasis. *J. Am. Med. Assoc.* 163:1248–51

50. Adler, R. H. 1967. A rebreather for prophylaxis and treatment of postoperative respiratory complications. *Dis. Chest* 52:640–47

51. Mahoney, P. D., Colgan, F. J. 1969. The effects of dead space rebreathing in the treatment of atelectasis. *Surg. Gynecol. Obstet.* 128:1283–88

52. Rudy, N. E., Crepeau, J. 1958. Role of intermittent positive pressure breathing postoperatively. *J. Am. Med. Assoc.* 167:1093–96

53. Becker, A., Barak, S., Braun, E., Meyers, M. P. 1960. The treatment of postoperative pulmonary atelectasis with intermittent positive pressure breathing. *Surg. Gynecol. Obstet.* 111:517–22

54. Sands, J. H., Cypert, C., Armstrong, R., Ching, S., Trainer, D., Quinn, W., Stewart, D. 1961. A controlled study using routine intermittent positive pressure breathing in the postsurgical patient. *Dis. Chest* 40:128–33

55. Baxter, W., Levine, R. S. 1969. An evaluation of intermittent positive pressure breathing in the prevention of postoperative pulmonary complications. *Arch. Surg.* 98:795–98

56. Anderson, W. H., Dossett, B. E. Jr., Hamilton, G. L. 1963. Prevention of postoperative pulmonary complications. Use of isoproterenol and intermittent positive pressure breathing on inspiration. *J. Am. Med. Assoc.* 186:763–66

57. Wilson, J. E. III, Pierce, A. K., Johnson, R. L. Jr., Winga, E. R., Harrell, W. R., Curry, G. C., Mullins, C. B. 1971. Hypoxemia in pulmonary embolism, a clinical study. *J. Clin. Invest.* 50:481–91

58. Bartlett, R. H., Krop, P., Hanson, E. L., Moore, F. D. 1970. Physiology of yawning and its application to postoperative care. *Surg. Forum* 21:222–75

59. Bartlett, R. H., Gazzaniga, A. B., Geraghty, T. 1971. The yawn maneuver: Prevention and treatment of postoperative pulmonary complications. *Surg. Forum* 22:196–98

60. Van de Water, J., Watring, W. G., Linton, L. A., Murphy, M., Byron, R. L. 1972. Prevention of postoperative pulmonary complications. *Surg. Gynecol. Obstet.* 135:229–33

61. Radigan, L. R., King, R. D. 1960. A technique for the prevention of postoperative atelectasis. *Surgery* 47:184–87

62. McCabe, R., Reid, W. M., Knox, W. G. 1962. Evaluation of the use of a tempo-

rary percutaneous endotracheal catheter in the treatment and prevention of postoperative pulmonary complications. *Ann. Surg.* 156:5–8
63. Sizer, J. S., Frederick, P. L., Osborne, M. P. 1966. The prevention of postoperative pulmonary complications by percutaneous endotracheal catheterization. *Surg. Gynecol. Obstet.* 123:336–40
64. Thoren, L. 1954. Post-operative pulmonary complications. Observations on their prevention by means of physiotherapy. *Acta Chir. Scand.* 107:193–205
65. Palmer, K. N. V., Sellick, B. A. 1953. The prevention of postoperative pulmonary atelectasis. *Lancet* 1:164–68
66. Stein, M., Cassara, E. L. 1970. Preoperative pulmonary evaluation and therapy for surgery patients. *J. Am. Med. Assoc.* 211:787–90
67. Cluver, G. H., Makel, H. P., Beecher, H. K. 1951. Frequency of aspiration of gastric contents by the lungs during anesthesia and surgery. *Ann. Surg.* 133:289–92
68. Berson, W., Adriani, J. 1954. "Silent" regurgitation and aspiration during anesthesia. *Anesthesiology* 15:644–49
69. Edwards, G., Morton, H. J. V., Pask, E. A., Wylie, W. D. 1956. Deaths associated with anaesthesia: a report on 1,000 cases. *Anesthesia* 11:194–220
70. Awe, W. C., Fletcher, W. S., Jacob, S. W. 1966. The pathophysiology of aspiration pneumonitis. *Surgery* 60:232–39
71. Kakkar, V. V., Field, E. S., Nicolaides, A. N., Flute, P. T., Wessler, S., Yin, E. T. 1971. Low doses of heparin in prevention of deep-vein thrombosis. *Lancet* 2:669–70
72. Kakkar, V. V., Corrigan, T., Spindler, J., Fossard, D. P., Flute, P. T., Crellin, R. Q., Wessler, S., Yin, E. T. 1972. Efficacy of low doses of heparin in prevention of deep-vein thrombosis after major surgery. *Lancet* 2:101–6
73. Clagett, G. P., Salzman, E. W. 1974. Prevention of venous thromboembolism in surgical patients. *N. Engl. J. Med.* 290:93–96
74. Sevitt, S. 1962. Venous thrombosis and pulmonary embolism: their prevention by oral anticoagulants. *Am. J. Med.* 33:703–16
75. Morrell, M. T., Dunnill, M. S. 1968. The post-mortem incidence of pulmonary embolism in a hospital population. *Br. J. Surg.* 55:347–52
76. Roberts, V. C., Cotton, L. T. 1974. Prevention of postoperative deep vein thrombosis in patients with malignant disease. *Br. Med. J.* 1:358–60
77. Cranley, J. J., Canos, A. J., Sull, W. J. 1976. The diagnosis of deep vein thrombosis. Fallibility of clinical symptoms and signs. *Arch. Surg.* 111:34–36
78. Barritt, D. W., Jordan, S. C. 1960. Anticoagulant drugs in the treatment of pulmonary embolism. A controlled trial. *Lancet* 1:1309–12
79. Coon, W. W., Willis, P. W. III, Symons, M. J. 1969. Assessment of anticoagulant treatment of venous thromboembolism. *Ann. Surg.* 170:559–67
80. Salzman, E. W., Harris, W. H., DeSanctis, R. W. 1971. Reduction in venous thromboembolism by agents affecting platelet function. *N. Engl. J. Med.* 284:1287–92
81. Husni, E. A., Ximenes, J. O. C., Goyette, E. M. 1970. Elastic support of the lower limbs in hospital patients. *J. Am. Med. Assoc.* 214:1456–62
82. Prevention of fatal postoperative pulmonary embolism by low doses of heparin. An International Multicentre Trial. 1975. *Lancet* 2:45–51
83. Gaensler, E. A., Weisel, R. D. 1973. The risks in abdominal and thoracic surgery in COPD. *Postgrad. Med.* 54:183–91
84. Miller, W. F., Wu, N., Johnson, R. L. Jr. 1956. Convenient method of evaluating pulmonary ventilatory function with single breath test. *Anesthesiology* 17:480–93
85. Ravin, M. B. 1971. Comparison of spinal and general anesthesia for lower abdominal surgery in patients with chronic obstructive pulmonary disease. *Anesthesiology* 35:319–22
86. Jones, A. M., Howitt, G. 1965. Eisenmenger syndrome in pregnancy. *Br. Med. J.* 1:1627–31
87. Skinner, J. F., Pearce, M. L. 1964. Surgical risk in the cardiac patient. *J. Chronic Dis.* 17:57–72

Ann. Rev. Med. 1977. 28:223–34
Copyright © 1977 by Annual Reviews Inc. All rights reserved

ACUPUNCTURE: CONTEXT AND CRITIQUE

❖7239

Bernard S. Millman, M.D.
Department of Anesthesia, Stanford University Medical School, Stanford, California 94305

> The variable composition of man's body hath made it an instrument easy to distemper. . . . The office of medicine is but to tune this curious harp of man's body and to reduce it to harmony.
>
> Francis Bacon (1)

INTRODUCTION

Orthodox medicine, like many of our institutions, is being challenged by seemingly new ideas disguised as strange cults (2–4). The introduction of a prescientific healing practice such as acupuncture into American life would seem ectopic were it not a thread in a metaphysical tapestry being woven in Western civilization for the past three decades. Its design was conceived in the despair of accepted social values promulgated by existential philosophy and manifested as the beatnik phenomenon of the 1950s. As existentialism developed into the hedonistic and idealistic youthful counterculture of the 1960s, the apparent irrelevance of external values led to a search that now is praised as the "new humanism" or condemned as the "new narcissism" (2, 5). This movement, gaining impetus from social and economic developments in the 1970s, is characterized by a return to folkways and an interest in mysticism. By eulogizing the irrational, the implication is made that man is greater than the sum of his parts. A sympathetic current in medicine evolving from psychosomatic research is the "holistic approach," which embodies the ancient principle that health represents a harmony between nature, society, and man, and reflects the unity of mind and body (6, 7). This concept is also the basis for acupuncture, yoga, and other forms of healing and is not new to Western medicine. In fact, it was the crux of a debate between Pidoux, who espoused the theory that the patient's constitution predisposes to a given disease, and Pasteur, who emphasized the role of specific agents as etiologic factors in disease (1).

Acupuncture enters the American medical literature as early as 1836 (8). It was even recommended by Osler for lumbago (9). While always extant among the Chinese community in the United States, its major impetus came from Reston's historic acupuncture treatment for postoperative pain in China in 1971 and former President Nixon's trip to China in 1972, which whetted the American enthusiasm for chinoiserie (10). The American medical community has not strongly resisted acupuncture when practiced in a safe manner on patients regarded as "hopeless" after orthodox treatments have failed. However, the official stand of organized medicine has been cautious, emphasizing the need for controlled studies (10). The proliferation of acupuncture newsletters, books, courses, materials, and journals (e.g. *American Journal of Chinese Medicine, American Journal of Acupuncture,* and *Acupuncture and Electro-Therapeutics Research*) has enhanced the enthusiasm for acupuncture's therapeutic merits and exploration of its scientific basis (4, 10).

Chinese medical philosophy, which includes herbology and exercises, as well as acupuncture, is a coherent and closed theory of treatment and diagnosis. As with Western medicine, theory probably developed from observations of patients and their environments. The theory was systematized and recorded by the eighth century BC, although its origins are attributed to sources as early as 2700 BC (11). Traditional acupuncture theory has many similarities to an electrical circuit in that there is a source of energy production called the "triple heater organ," which produces a finite amount of vital energy, called ch'i, during a lifetime. The vital energy flows along fixed paths called meridians, which are interconnected with each other and the organs of the body. Flow in these paths has a circadian rhythm, so that at various times and seasons it may normally vary. As with electrical energy, vital energy has a bipolar character, i.e. positive and negative, which are called "yang" and "yin," respectively. Ch'i within the body reflects the vitality of the universe and society. Therefore one would expect more illness during natural or social disasters.

In this system, health is defined as the state in which the flow of vital energy in the meridians is normal. Diagnosis in Chinese medicine is made on the basis of the patient's history of changes in behavior, appetite, and feelings. Physical examination consists of observation of the skin, eyes, face, tongue, and breath for color, consistency, and odor (10, 11). Special diagnostic procedures consist of (*a*) pulse diagnosis, during which deep and superficial palpations of three sites on each radial artery are made to evaluate the energy flow in each of the twelve meridians; (*b*) palpation of skin and subcutaneous tissues for areas of tenderness or numbness; and, more recently, (*c*) measurement of the electrical resistance of the skin with an ammeter at specific sites on the wrists and ankles that also relate to the flow of energy in the meridians (12–14). Therapy consists of the insertion of needles, the application of small electrical currents, or other manipulations at specific sites to adjust the energy flow to the normal distribution. The meridian system has not been demonstrated histologically, although it has been claimed that radionuclides injected into an acupuncture point will be detected along its respective meridian (15).

Ironically, traditional acupuncture is no longer extensively practiced in China, but is now practiced in the West, particularly in France, Germany, England, and the United States (16). This development is a consequence of Chairman Mao's directive to Chinese scientists to explore the "treasure-house" of Chinese medicine in 1954 (10). Therefore, much of ancient medical philosophy has been discarded in China as the result of current empirical data. Because of the great Proletarian Cultural Revolution of 1965–1966, traditional and Western medicine have been thoroughly integrated—perhaps more for political purposes than therapeutic benefit. In so doing, the revolution destroyed the elitist status of physicians and forced them to study and apply a uniquely Chinese system of healing that could be easily administered to a huge population.

Although acupuncture enjoyed a brief period of Western enthusiasm in the eighteenth century as the result of Jesuit missionary contact with the Orient, it was not popularized again until the 1920s, when Georges Soulié de Morant, a banker and diplomat to China, demonstrated the technique in France (16, 17). Modern acupuncture variants carry a theme from "man as a microcosm of the universe" to "parts of man as representations of the whole man," e.g. the ear has the outline of a fetus; therefore, points corresponding to fetal organs superimposed on the ear can be used to treat adults (18). Similar speculative analogies are the bases for schools of nose, hand, and scalp acupuncture (19). More scientific apologists invoke the proximity of the central nervous system (CNS) projections of the acupuncture stimulation site and the pain path as the rationale for the selection of treatment loci (20).

PHENOMENOLOGY OF ACUPUNCTURE

Acupuncture Points (Loci)

Traditionally, these points have been located by palpation as interfascicular muscular depressions or in relation to other anatomic landmarks (13). They may also be located with an ammeter as areas of low electrical resistance, although not all low-resistance loci are acupuncture points (14, 21). There is a high correlation of acupuncture points with nerve trunks, motor points, and trigger points (22–24). Some experimental work suggests that many points have specific effects (25, 26). In clinical practice, however, there is controversy as to whether accurate needle placement affects outcome when used to treat disease (20, 27).

Te Ch'i

When needles are accurately placed, patients will comment that they feel a numbness, heaviness, ache, or paresthesia at the site of insertion or in the diseased part (13). This phenomenon is called te ch'i ("summoning the vital energy"). Since many of the acupuncture points are related to neural structures, such sensations are easily explained. Needles are commonly placed in the nerve proximal to the diseased area in that nerve's distribution (13, 22). The feeling of te ch'i also correlates with the appearance of muscle potentials on the oscilloscope. These potentials may be a reflex phenomenon, as they are abolished by spinal anesthesia. Apparently, perception of this sensation is mediated by the lateral spinothalamic tracts of the spinal cord (22, 28). Te ch'i is essential for acupuncture analgesia and is abolished by infiltrating the acupuncture point by local analgesics (29). Often, no effort is made to elicit this sensation in animals, although therapeutic and analgesic effects are still noted (26). The observation that te ch'i can be elicited anywhere by persistent needling and has many qualities of deep pain sensation as well as paresthesia raises the question of whether this is a unique feature of acupuncture points, although it may be necessary for specific purposes (28, 30). It seems likely that te ch'i at least serves to convince the patient and practitioner that something is happening (31).

ACUPUNCTURE MECHANISMS: AN ECLECTIC PERSPECTIVE

Traditional Chinese medical philosophy is attractive as a theory representing a distillate of much experience and providing a coherent model for practice. However, literal anatomic documentation of the meridian system is reported by only one investigator (15). While each of the following alternative mechanisms of acupuncture was also presented as a unifying model, it is apparent that none is. It is more likely that acupuncture represents many phenomena, much as the ingestion of aspirin for headache has both antiphlogistic and reassuring aspects (4). This view is especially attractive in light of the diverse effects and structures associated with acupuncture points.

Humoral-Biochemical Mechanism

This speculative mechanism is based on the observations of histamine release at the acupuncture site, a phenomenon akin to dermatographia, and the postulated hormonal effects of tissue trauma caused by needling (20, 27, 32). Histamine, bradykinin, and other hormones released by tissue trauma have many effects on microcirculation, muscle tissue, heart, reticuloendothelial system, etc. Physical stimulation may also result in steroid secretion. This hypothesis implies that accurate needle placement is not essential to therapy and explains why novices are so successful as therapists (20, 27). There is little well-controlled experimental work in this area, but it seems unlikely that acupuncture produces enough trauma or hormone release to accomplish all of its wondrous therapeutic claims on a direct peripheral hormonal basis. Vascular occlusion of the acupunctured extremity does not impair its analgesic effect. This would suggest a neural or psychologic explanation at least for analgesia, if not chronic disease (33).

Neural Mechanisms

The phenomena of acupuncture points and te ch'i are especially suggestive of a neural mechanism. The gate theory of Melzack & Wall is commonly invoked to explain acupuncture analgesia. This theory proposes that a physiologic "gate" in the substantia gelatinosia inhibits peripheral small-fiber input associated with pain when there is concurrent large-fiber stimulation, e.g. from pressure. This theory derives support from the observation that acupuncture analgesia is most effective when done in the same dermatome as the surgical field (34–36). Because of its spectacular nature, a great deal of attention has been drawn to this form of surgical analgesia. The myriads of cases reported from China attest to its effectiveness (35, 37, 38).

Claims that acupuncture analgesia can be done in infants and even animals are particularly disarming. It seems likely that reassurance, rehearsals, premedication, deft surgeons, and supplementation with local and even general analgesia contribute to its effectiveness in humans (10). In fact, the latest acupunctural anesthetic feat has been to reduce the number of needles from 80 to 1, if not to dispense with the needles entirely (39). Its limitations and the rituals leading up to surgery are reminiscent of hypnosis (31). The gate theory itself is a subject of controversy, as its internuncial system has not been documented anatomically and the theory does not explain why painful stimulation at one site decreases susceptibility to pain at others (40, 41). Acupuncture's effects in fact seem to be related to small-fiber stimulation (22, 28). Perhaps its mechanism is related to hyperstimulation analgesia in that acupuncture may be a minor noxious stimulus that confers analgesia by competing with other noxious stimuli (33, 41).

There are many limitations to acupuncture analgesia. The long induction times (20–30 min), and lengthy rehearsals and explanations to the patient of the events that will occur during surgery preclude its use for emergencies (39). It is not recommended for the "mentally disturbed, nervous, low pain threshhold, or unwilling" patient. Abdominal relaxation is poor and this form of analgesia requires an adaptable surgeon who works rapidly, but carefully (35).

Animal studies indicate inhibition by acupuncture of thalamic neuronal discharges associated with noxious stimulation (42–44). The duration of this effect is too brief to be clinically meaningful in chronic pain states. Most astonishing are the controlled observations that analgesia may be transferred by cerebrospinal fluid from acupunctured to nonacupunctured animals, and that antagonism of acupuncture analgesia results from intraventricular injections of atropine (44)! A reflex mechanism has also been postulated for acupuncture, since the management of chronic pain problems with acupuncture has much in common with trigger-point injections of local anesthetics for myofascial syndromes (45). Here, too, is much controversy, for as our knowledge of chronic pain problems improves, it has become apparent that we have many treatments for which there are no diseases, e.g. for neuralgia, rheumatism, and lumbago (46). These diagnoses confer pseudo-omniscience on the physician and distract the physician and patient from a more relevant psychiatric diagnosis. The lack of a peripheral "disease" in some of these patients can be demonstrated by the persistence of the pain complaint despite surgical analgesia by nerve block of the affected area (47).

Bioelectric Mechanism

The electrical detection of acupuncture points by probes and Kirlian photography; the demonstration of consistent changes in skin conductivity with disease; and the regeneration of limbs in amputated animals placed in an electric field are exciting food for a feast of speculation (14, 15, 21, 48, 49). One bioelectric school of acupuncture attributes electrical changes to autonomic function and attempts to regulate this division of the nervous system by the application of current (14). However, correlative studies of consistent changes in conductivity at specific acupuncture points with specific disease entities are not available. The application of electrical stimulation seems as effective as other types of stimulation, e.g. massage, ultrasound, laser, vibration, or needling (50). Except for needling, none of these types of stimulation require the te ch'i phenomenon. The wide variety of these allegedly effective types of stimulation again makes one question the role of acupuncture loci.

Psychologic Mechanisms

Because the pathophysiology of disease and surgery and the pharmacology of drugs are relatively easy to observe in a scientific fashion, one may be lured into the delusion that these are the stuff of clinical practice. As laboratory creations, these lose meaning when dealing with chronic illness, which has personal, social, and cultural matrices (51, 52). Many of the so-called hopeless benign, functional conditions, particularly chronic pain, represent symptoms of anxiety and/or depression and, less commonly, hallucinosis, delusion, obsessive or phobic behavior, malingering, distraction, addiction, boredom, existential need, communication inadequacy, and habit (53–57). Furthermore, organic illness does not preclude these behaviors from complicating management (58). Patients with these problems become frustrating when their problem continues to be refractory to well-intentioned inappropriate treatment. Psychologic tests indicate hysteric, obsessive-compulsive, inadequate, or schizoid personalities clinically manifested as dependency. A benevolent acupuncturist has much to offer in terms of a sympathetic ear, patience, a willingness to ascribe a physical basis to an emotional disorder, an exciting and mysterious modality, and a chance for the patient to "get even" with all those doctors who implied a psychiatric diagnosis (59–64). However, the polysurgery patient is difficult to cure even for the acupuncturist (65). This difficulty is attributed to surgical interruption of the meridians, much as persistent pain after laminectomy is attributed to adhesions. At present, empathy and perhaps behavior modification may be the most relevant tools to manage the complaints of chronic illness.

PLACEBO EFFECTS Most uncontrolled clinical studies of new modalities report improvement in 80–90% of patients (66, 67). In some controlled studies, the placebo effect may be similarly high (67)! This phenomenon seems especially operative when new therapies are introduced, at which time the therapist's and patient's enthusiasm play the major roles in efficacy (68). The variables contributing to the placebo effect can produce remarkable responses, e.g. paradoxical drug effects, miraculous cures, toxicity, and even death (69–72). When nonspecific factors are controlled, the placebo is usually half as effective as the active agent and provides satisfactory relief in one third of the patients in a wide variety of conditions over a long period of time, e.g. angina, rhinitis (71). Acupuncture, and indeed many respected, radical forms of "scientific" therapy for chronic pain, e.g. rhizotomy or cordotomy, have statistical patterns of efficacy suggestive of placebo (73–79). While statistical patterns may be similar to placebo, this may only be coincidence. What is of major importance is that less dangerous and less costly methods may be available to produce similar effects. With long-term follow-up it seems likely that the efficacy of acupuncture and "radical" Western treatments is a victim of the time course of chronic diseases, especially if these are subject to exacerbations and remissions associated with psychic distress (80, 81).

Both placebo and nonplacebo reactors have stereotyped responses to psychologic tests (67). Notably, the placebo reactor has a high level of chronic anxiety (61, 82). The placebo effect is most marked when the patient is under stress (83). Of special interest is that the route or style of placebo administration also has a role in effectiveness (84). The placebo effect is not purely psychological, but represents a general term, including many factors over which the therapist has no control and which contribute to producing a given response.

HYPNOSIS/SUGGESTION Much of whether acupuncture is hypnosis depends on whether one believes in this unique psychic state. The external manifestations of hypnosis are culturally dependent, e.g. trance versus seizure. There is a wealth of evidence suggesting no difference between hypnosis and consciousness and a contrary literature suggesting at least some unusual behaviors characteristic of hypnosis (85, 86). However, when an effort is made to control for the hypnosis variable, it can be distinguished from other placebo effects (87). Hypnoanesthesia for surgery may be accomplished without a deep trance (31, 88). Indeed, there is a correlation between hypnotic susceptibility and the exhibition of acupuncture analgesia (89, 90). Contrarily, acupuncture analgesia may be demonstrated in poor hypnotic subjects. Despite the hundreds of thousands of successful reports of acupuncture analgesia, the vast majority of Chinese patients elect conventional anesthetic techniques (10). Acupuncture is not recommended for infants without heavy sedation (10, 35). Children, incidentally, can best be hypnotized after the age of five (86).

Animal studies of acupuncture may be irrelevant to human acupuncture analgesia, as many species become cataleptic when restrained, e.g. the rabbit, coyote, octopus, rat, and crocodile (91, 92). Repetitive stimuli and pressures on body parts, which are features of acupuncture, are also adequate induction stimuli for animal "hypnosis" (26, 44, 91). Novelty of stimulation is a particularly potent induction factor. During the cataleptic state, animals do not respond to painful manipulations. It is postulated that the tonic immobility response to restraint is an unlearned fear reaction (91). The distribution of analgesia is not reported in animal catalepsy, but seems specific in the case of acupuncture, with location dependent on stimulation site (26). EEG studies of immobilized animals have patterns that are distinct from those of human hypnotic subjects (91–93). Comparisons of human EEGs under acupuncture and under hypnosis are controversial, as the characteristic EEG pattern under hypnosis is uncertain.

STATUS OF NONCLINICAL INVESTIGATIONS

Using a double-blind technique, acupuncture analgesia can be distinguished from placebo effects. This suggests a physiologic basis (94). On the other hand, it can be shown that acupuncture changes the subject's criteria for reporting pain (95). As is commonly the problem with such studies, the magnitude or duration may be statistically significant, but clinically unimportant (96). Similarly, animal studies are fascinating, sometimes illuminating, but often of questionable significance in the study of human subjective phenomena (97, 98).

STATUS OF CLINICAL INVESTIGATIONS

Uncontrolled clinical studies in more than 20,000 patients at UCLA suggest that acupuncture is most useful for migraine and tension headaches, osteoarthritic pain, bursitis, tenosynovitis, and other musculoskeletal problems. There have been no serious complications in this extensive experience (D. E. Bresler, R. J. Kroening, M. P. Volen, submitted for publication). Controlled studies have been hampered by (*a*) the lack of qualified therapists, (*b*) difficulty in double-blind design when a physical modality is used, and (*c*) a poor understanding of what constitutes an adequate trial of acupuncture therapy. The proliferation of modalities, the varieties of acupuncture styles, and the departures from classical acupuncture make it difficult to compare data from one clinical study to others or to comment on the efficacy of acupuncture in general. Acupuncture journals even carry articles on nerve blocks and trigger-point injections with local anesthetics. The use of electrical stimulation through acupuncture needles has invited comparisons with transcutaneous nerve stimulation. The latter technique has even been advocated at acupuncture points (P. Raj, submitted for publication). It seems likely that there will be no "modern acupuncture," but a divergence of techniques, each having a unique mechanism. The validity of each will have to be established separately.

INEXPLICABLE FINDINGS

A recent double-blind, controlled, crossover study of acupuncture in asthmatics utilized changes in pulmonary-function studies as a measure of effectiveness in reversing methacholine-induced bronchospasm. Acupuncture was less effective than isoproterenol, but more effective than placebo (D. P. Tashkin, D. E. Bresler, R. J. Kroening, H. Kerschnar, R. L. Katz, A. Coulson, submitted for publication). Other studies imply a variety of objective changes attributed to specific acupuncture points, such as increased mesenteric circulation, immunologic effects, and predictable zones of analgesia. These results are not easily explained, widely documented, or of clinically significant magnitude, but indicate areas for future physiologic research (20, 25, 26, 99–101). Uncontrolled observations in animals with chronic disease also suggest that acupuncture is not humbug, but is useful in managing emphysema, arthritis, laminitis, hepatitis, facial palsy, hip dysplasia, paralysis, herniated nucleus pulposus, and cardiac-valvular insufficiency in horses and dogs (102, 103).

An interesting reflection on acupuncture's mysterious cures and physiologic effects is that there are no similarly documented mysterious complications of its practice, although abortion is said to follow needling in the first metacarpal space, and convulsions occur in patients who ingest alcohol prior to treatment. Its benign complications, e.g. fatigue and exacerbation of symptoms, are those commonly seen with placebo treatments and its documented severe complications are generally the result of anatomic trespass (71, 104). This benignity may relate to 5000 years of evolving perfection, but is more reminiscent of George Bernard Shaw's observation of blasphemy at Lourdes by the visual testimony of piles of "crutches and wheelchairs, but no wooden legs or glass eyes—or toupees," thereby hinting at a limitation of God's powers (6).

A ROLE FOR ACUPUNCTURE?

The relative safety of acupuncture makes it an attractive alternative to surgery and medications, with their attendant morbidity, mortality, and futility (6, 46, 66, 71, 105, 106). Its developing indication seems to be for functional diseases (22, 107). This recommendation may be premature and shallow, as there have been no long-term studies showing that patients "addicted" to surgery and medication will be content with innocuous procedures (106). What may be needed for the management of these problems is a courageous public and profession that recognize (*a*) the distinction between mind and body as a linguistic artifact, (*b*) the unsuitability of drugs and surgery for the treatment of occult interpersonal and intrapsychic problems, (*c*) the limitations of the "illness model" when applied to the extremes of behavior or physiologic measurement, and (*d*) the resources of the individual that might be enhanced for health (6, 46, 108–111). Some patients will fail to comprehend these possibilities so that cult medicine will always be present in a free society as a sociologic phenomenon akin to patent medicines (60, 65). However, one need not endorse cultism as an excuse for therapeutic nihilism.

Preliminary studies suggest that acupuncture is more than superstition. However, the susceptibility of humans to psychologic manipulation has been a confounding variable in acupuncture studies as well as an important adjunct to the management of human illness. Scientific reports on the efficacy of acupuncture for chronic disease are too few to enable a reasoned opinion as to its place in medicine even if a foundation in physiology or psychology can be established. Perhaps the greatest contribution it will make to scientific medicine is to give the latter an opportunity to examine itself by comparison, and with equally stringent standards. As the great "treasure house" of science is searched, we shall discover that rationalism is also a humanism (5).

ACKNOWLEDGMENTS

The author wishes to express his appreciation to Ellis N. Cohen, M.D. and David E. Bresler, Ph.D. for their constructive suggestions during the preparation of this paper.

Literature Cited

1. Dubos, R. J. 1959. The gold-headed cane in the laboratory. *Public Health Rep.* 69:365–71
2. Marin, P. 1975. The new narcissism. *Harper's.* October: 45–56
3. Glasser, W. 1972. The civilized identity society. *Saturday Rev.* 55:26–31
4. Davis, D. L. 1975. The history and sociology of the scientific study of acupuncture. *Am. J. Chin. Med.* 3:5–26
5. Sartre, J. 1956. Existentialism is a humanism. In *Existentialism from Dostoevsky to Sartre,* ed. W. Kaufmann, pp. 287–311. Cleveland: Meridian. 321 pp.
6. Frank, J. D. 1975. The faith that heals. *Johns Hopkins Med. J.* 137:127–31
7. Fabrega, H. Jr. 1974. *Disease and Social Behavior: An Interdisciplinary Perspective,* pp. 138–141. Cambridge, Mass: MIT Press. 341 pp.
8. Lee, W. M. 1836. Acupuncture as a remedy for rheumatism. *South. Med. J.* 1:129–33
9. Riddle, J. W. 1974. Report of the New York State Commission on Acupuncture. *Am. J. Chin. Med.* 2:289–318
10. Bonica, J. J. 1974. Therapeutic acupuncture in the People's Republic of

China. *J. Am. Med. Assoc.* 228:1544–51

11. Kao, F. F. 1973. China, Chinese medicine and the Chinese medical system. *Am. J. Chin. Med.* 1:1–59
12. Pennell, R. J., Heusner, G. D. 1973. How to diagnose the Chinese pulses. *Am. J. Acupunct.* 1:168–73
13. Tan, L. T., Tan, M., Veith, I. 1973. *Acupuncture Therapy—Current Chinese Practice.* Philadelphia: Temple Univ. Press. 159 pp.
14. Sato, T., Nakatani, Y. 1974. Acupuncture for chronic pain in Japan. *Adv. Neurol.* 4:813–18
15. Tiller, W. A. 1973. Some energy field observations of man and nature. In *Galaxies of Life,* ed. S. Krippner, pp. 71–112. New York: Gordon & Breach. 182 pp.
16. Wild, R. A. 1973. Acupuncture—A current legal analysis. *Am. J. Chin. Med.* 1:123–33
17. Simpson, E. 1972. Acupuncture (à la francaise). See Ref. 3, pp. 47–49
18. Chapman, C. R. 1974. Acupuncture: Some considerations for the control of pain in dentistry. *J. Prosthet. Dent.* 31:441–51
19. Liu, T. H., Sadove, M. S. 1974. Scalp needle therapy: Acupuncture treatment for central nervous system disorders. *Am. J. Chin. Med.* 2:261–69
20. Bresler, D. E., Kroening, R. S. 1976. Three essential factors in effective acupuncture therapy. *Am. J. Chin. Med.* 4:81–86
21. Brown, M. L., Ulett, G. A., Stern, J. A. 1974. Acupuncture loci: Techniques for location. *Am. J. Chin. Med.* 2:67–74
22. Matsumoto, T. 1974. *Acupuncture for Physicians,* pp. 140–150. Springfield: Thomas. 224 pp.
23. Gunn, C. C., Ditchborn, F. G., King, M. H., Renwick, G. J. 1976. Acupuncture loci: A proposal for their classification according to their relationship to known neural structures. *Am. J. Chin. Med.* 4:183–95
24. Lawenschuss, O. 1975. Electroacupuncture progress and effectiveness. *Am. J. Acupunct.* 3:347–51
25. Matsumoto, T., Hayes, M. F. 1973. Acupuncture, electric phenomenon of the skin, and postvagotomy intestinal atony. *Am. J. Surg.* 125:176–80
26. Matsumoto, T., Ambruso, V., Levy, B. A., Bednanek, J. 1974. Acupuncture analgesia in animals: Study of specific location(s) of acupuncture points. *Am. Surg.* 40:340–44
27. Platt, H. V. 1974. The trouble with acupuncture. *Am. J. Acupunct.* 2:73–76
28. Small, T. J. 1974. The neurophysiological basis for acupuncture. *Am. J. Acupunct.* 2:77–87
29. Lowe, W. C. 1973. *Introduction to Acupuncture Anesthesia.* Flushing, NY: Med. Exam. Publ. 101 pp.
30. Kellgren, J. T. 1949. Deep pain sensibility. *Lancet* 1:943–49
31. Kroger, W. S. 1973. The scientific rationale for acupuncture analgesia. *Psychosomatics* 14:191–94
32. Popkin, R. S. 1972. Histamine explanation of acupuncture. *J. Am. Med. Assoc.* 220:1359
33. Wancara, I., König, G. 1974. On the neurophysiologic explanation of acupuncture analgesia. *Am. J. Chin. Med.* 2:193–98
34. Melzack, R., Jeans, M. 1974. Acupuncture analgesia—a psychophysiological explanation. *Minn. Med.* 57:161–66
35. Small, T. J. 1974. Acupuncture anesthesia: A review. *Am. J. Acupunct.* 2:147–63
36. Staff of PLA Hospital 1973. Some insights concerning the principles of acupuncture anesthesia. *Am. J. Chin. Med.* 1:167–71
37. Department of Surgery, Peking Children's Hospital 1973. Acupuncture anesthesia in pediatric surgery: Report of 1,308 cases. *Chin. Med. J.* 2:23
38. Department of Surgery, Tumor Institute of the Chinese Academy of Medical Sciences 1973. Study of 1,230 surgical cases of carcinoma of esophagus and gastric cardia. *Chin. Med. J.* 2:28–29
39. Spoerel, W. E. 1975. Acupuncture analgesia in China. *Am. J. Chin. Med.* 3:359–68
40. Cervero, F., Iggo, A., Ogawa, H. 1976. Nociceptor driven dorsal horn neurones in the lumbar spinal cord of the cat. *Pain* 2:5–24
41. Melzack, R. 1975. Prolonged relief of pain by brief, intense transcutaneous somatic stimulation. *Pain* 1:357–73
42. Chang, H. T. 1974. Integrative action of thalamus in the process of acupuncture for analgesia. *Am. J. Chin. Med.* 2:1–39
43. Linzer, M., Van Atta, L. 1974. Effects of acupuncture stimulation on activity of single thalamic neurons in the cat. *Adv. Neurol.* 4:799–801
44. Research Group of Acupuncture Anaesthesia, Peking Medical College 1974. The role of some neurotransmitters of brain in finger-acupuncture analgesia. *Sci. Sinica* 27:112–30

45. Frost, E. A. M., Hsu, C. Y. 1975. Neurophysiologic pathways in acupuncture. *Am. J. Acupunct.* 3:331–34
46. Meador, C. K. 1965. The art and science of non-disease. *N. Engl. J. Med.* 272:92–95
47. Winnie, A. P., Collins, V. J. 1968. The pain clinic. I. Differential neural blockade in pain syndromes of questionable etiology. *Med. Clin. North Am.* 52: 123–29
48. Fries, J. F., Richter, C. P. 1964. Lung cancer: detection by use of electrical skin resistance method. *Arch. Intern. Med.* 113:624–34
49. Becker, R. O. 1972. Augmentation and regenerative healing in man. *Clin. Orthop. Relat. Res.* 83:255–62
50. Khoe, W. H. 1974. The many modalities of modern acupuncture. *Am. J. Acupunct.* 2:272–74
51. Beecher, H. K. 1956. Limiting factors in experimental pain. *J. Chronic. Dis.* 4:11–21
52. Rynearson, E. K. 1973. The dying patient. *Bull. Mason Clin.* 27:123–31
53. Sternbach, R. S. 1974. Pain and depression. In *Somatic Manifestations of Depressive Disorders.* ed. A. Kiev, pp. 101–9. Amsterdam: Excerpta Medica. 128 pp.
54. Pedder, J. R. 1969. Psychosomatic disorders and psychosis. *J. Psychosom. Res.* 13:339–46
55. Szasz, T. S. 1968. The psychology of persistent pain. In *Pain,* ed. A. Soularic, J. Cohn, J. Charpentier, pp. 93–112. London & New York: Academic. 562 pp.
56. Parsons, T. 1951. Illness and the role of the physician: A sociological perspective. *Am. J. Orthopsychiatry* 21:452–60
57. Fordyce, W. E. 1974. Office management of chronic pain. *Minn. Med.* 57:185–88
58. Woodeforde, J. M., Fielding, J. R. 1970. Pain and cancer. *J. Psychosom. Med.* 14:365–70
59. Delessio, D. J. 1974. Studies on pain and the Hawthorne effect. *Headache* 14:109–10
60. Snow, L. F. 1974. Folk medical beliefs and their implications for care of patients. *Ann. Intern. Med.* 81:82–96
61. Evans, F. J. 1974. The placebo response in pain reduction. *Adv. Neurol.* 4:289–300
62. Frank, J. D. 1974. Psychotherapy: The restoration of morale. *Am. J. Psychiatry* 131:271–74
63. Firman, G. J., Goldstein, M. S. 1975. The future of chiropractic: A psychosocial view. *N. Engl. J. Med.* 193:639–42
64. Sternbach, R. A. 1974. Varieties of pain games. *Adv. Neurol.* 4:423–40
65. Leung, S. J. 1973. Acupuncture treatment for pain syndrome. Part I—Treatment of sciatica. *Am. J. Chin. Med.* 1:317–26
66. Beecher, H. K. 1965. Surgery as placebo. A quantitative study in bias. *J. Am. Med. Assoc.* 176:1102–7
67. Lasagna, L., Mosteller, F., Von Felsinger, J. M., Beecher, H. K. 1954. A study of the placebo response. *Am. J. Med.* 16:770–79
68. Shapiro, A. K. 1960. A contribution to the history of the placebo effect. *Hum. Behav.* 5:109–35
69. Wolf, S. 1950. Effects of suggestion and conditioning on the action of chemical agents in human subjects. *J. Clin. Invest.* 29:100–9
70. Caroline, N. L., Schwartz, H. 1975. Chicken soup rebound and relapse of pneumonia. *Chest* 67:215–16
71. Beecher, H. K. 1959. *Measurement of Subjective Responses,* pp. 66–67. New York: Oxford Univ. Press. 494 pp.
72. Beecher, H. K. 1972. The placebo effect as a nonspecific force surrounding disease. In *Pain—Basic Principles, Pharmacology Therapy,* ed. J. P. Payne, R. A. P. Burt, pp. 175–80. Baltimore: Williams & Wilkins. 464 pp.
73. Kajdos, V. 1973. The treatment of allergic rhinitis with acupuncture *Am. J. Acupunct.* 1:52–55
74. Modell, J. R. 1975. Acupuncture for sensorineural hearing loss. *Arch. Otolaryngol.* 101:441–45
75. Kajdos, V. 1973. Acupuncture therapy of cervical syndrome. *Am. J. Acupunct.* 1:8–13
76. Man, P. L., Chen, A. 1974. Two year acupuncture follow-up study on 182 chronic pain patients. *Am. J. Acupunct.* 3:143–46
77. Gaw, A. C., Chang, L. W., Shaw, L. C. 1975. Efficacy of acupuncture on osteoarthritic pain—A controlled, double-blind study. *N. Engl. J. Med.* 293:375–78
78. Lee, P. K., Anderson, T. W., Modell, J. H., Saga, S. A. 1975. Treatment of chronic pain with acupuncture. *N. Engl. J. Med.* 232:1133–35
79. Rosomoff, H. L. 1969. Neurosurgical control of pain. *Ann. Rev. Med.* 20:189–200

80. Fordyce, W. E. 1976. *Behavioral Methods for Chronic Pain and Illness* St. Louis: Mosby. 236 pp.
81. Lane, F. M. 1973. Mental mechanisms and the pain of angina pectoris. *Am. Heart J.* 79:563–68
82. Chapman, C. R., Feather, B. W. 1973. Effects of diazepam on human pain tolerance and pain sensitivity. *Psychosom. Med.* 35:330–40
83. Beecher, H. K. 1956. Evidence for increased effectiveness of placebos with increased stress. *Am. J. Physiol.* 187: 163–69
84. Blackwell, B., Bloomfeld, S. S., Buncher, C. R. 1972. Demonstration to medical students of placebo responses and non-drug factors. *Lancet* 1: 1279–82
85. Barber, T. X. 1961. Physiologic effects of "hypnosis." *Psychol. Bull.* 58:390–419
86. Hilgard, E. R., Hilgard, J. R. 1975. *Hypnosis in the Relief of Pain.* Los Altos, Calif: Kaufmann. 262 pp.
87. McGlashan, T. H., Evans, F. J., Orne, M. T. 1969. The nature of hypnotic analgesia and placebo response to experimental pain. *Psychosom. Med.* 31: 227–46
88. Chisholm, N. A. 1972. Acupuncture analgesia. *Lancet* 1:540
89. Katz, R. L., Kao, C. Y., Spiegel, H., Katz, G. J. 1974. Pain, acupuncture, hypnosis. *Adv. Neurol.* 4:819–25
90. Moore, M. E., Berk, S. N. 1975. Acupuncture, placebo, and hypnotic susceptibility. *Arthritis Rheum.* 18:416–17
91. Ratner, S. C. 1967. Comparative aspects of hypnosis. In *Handbook of Clinical and Experimental Hypnosis,* ed. J. E. Gordon, pp. 550–87. New York: MacMillan. 653 pp.
92. Parson, S. L. C., Huggins, S. E. 1965. Effects of temperature on electroencephalogram of the caiman. *Proc. Soc. Exp. Biol. Med.* 120:422–26
93. Ulett, G. A., Akpinav, S., Ith, T. M. 1972. Quantitative EEG analysis during hypnosis. *Electroencephalogr. Clin. Neurophysiol.* 33:361–68
94. Anderson, D. G., Jamieson, J. L., Man, S. C. 1974. Analgesic effects of acupuncture on the pain of ice water. A double-blind study. *Can. J. Psychol.* 28:239–44
95. Clark, W. C., Young, J. C. 1974. Acupuncture analgesia? Evaluation by signal detection theory. *Science* 184: 1096–97
96. Chapman, C. R. 1975. Psychophysiological evaluation of acupunctural analgesia: Some issues and considerations. *Anesthesiology* 43:501–6
97. Millman, B. S. 1974. Investigating acupuncture. *Anesthesiology* 40:611–12
98. Chan, S. H. H. 1975. Suppression of a polysynaptic reflex by acupuncture and a possible underlying presynaptic mechanism in the spinal cord of the cat. *Exp. Neurol.* 48:336–42
99. Kvirchishvili, V. I. 1974. Projections of different parts of the body on the surface of the concha auriculae in humans and animals. *Am. J. Acupunct.* 2:208
100. Nemerof, H., Rothman, I. 1974. Acupuncture and hypnotism: Preliminary experiments and a warning. *Am. J. Clin. Hypn.* 16:156–59
101. Lee, G. T. C. 1974. A study of electrical stimulation of acupuncture locus Tsusanli (St-36) on mesenteric circulation. *Am. J. Chin. Med.* 2:53–56
102. DeGroot, A., Bresler, D. E. 1974. Acupuncture—a pilot program in horses. *J. Am. Vet. Med. Assoc.* 164:367–68
103. 1973. Veterinary acupuncture: Old Wives' tale or new clinical tool? *Mod. Vet. Pract.* 54:37–42
104. Peacher, W. G. 1975. Adverse reactions, contraindications and complications of acupuncture and moxibustion. *Am. J. Chin. Med.* 3:35–46
105. Handler, S. 1971. Bring back the mustard plaster. *Minn. Med.* 54:973–79
106. Spiro, H. R. 1968. Chronic factitious illness—Munchausen's syndrome. *Arch. Gen. Psychiatry* 18:569–79
107. Chapman, C. R. 1975. The management of chronic pain in the United States and Japan: A cross cultural comparison. *Clin. Med.* 82:39–47
108. Graham, D. T. 1967. Health, disease, and the mind-body problem: Linguistic parallelism. *Psychosom. Med.* 29:52–71
109. Szasz, T. S. 1961. *The Myth of Mental Illness.* New York: Harper & Row. 337 pp.
110. Spiro, H. M. 1976. Pain and perfectionism—the physician and the "pain patient." *N. Engl. J. Med.* 294:829–30
111. Miller, S., Remen, N., Barbour, A., Nakles, M. A., Miller, S., Garell, D. 1975. *Dimensions of Humanistic Medicine.* San Francisco: Inst. Study Humanistic Med. 101 pp.

Ann. Rev. Med. 1977. 28:235–50
Copyright © 1977 by Annual Reviews Inc. All rights reserved

HYPERTROPHIC CARDIOMYOPATHY

❖7240

Pravin M. Shah, M.D.
Department of Medicine, Cardiology Unit, University of Rochester School of Medicine and Dentistry, Rochester, New York 14642

DEFINITION

The term *hypertrophic cardiomyopathy* (HCM) refers to "idiopathic" or "primary" cardiac hypertrophy characterized by considerable increase in ventricular muscle mass without associated increase in cavity size. Implicit in this definition is a possibility of common pathological endpoint of etiologically diverse disorders. Since nearly 90% of patients with HCM have the characteristic bizarre asymmetric hypertrophy of the heart with or without outflow-tract obstruction and form a characteristic clinical syndrome of a familial disorder, this form is especially emphasized here. This disorder has been variously labeled as hypertrophic obstructive cardiomyopathy (HOCM), idiopathic hypertrophic subaortic stenosis (IHSS) and muscular subaortic stenosis (MSS).

HISTORICAL PERSPECTIVES

There has been, through the years, a considerable interest in affections of the myocardium where no valvular lesion was evident at autopsy. In man's need to attribute rational explanations to unknown natural phenonomena, these disorders were presumptively labeled in the latter part of the 19th and early 20th century as "burn-out hypertensive heart disease," largely due to association of massive hypertrophy in patients known to have severe progressive systemic hypertension. It is probably in this "cloak" of diagnosis that the HCM hearts of yesteryears rest in the pathological archives of Europe. Refinement of bedside diagnosis of heart disease by Wood and development of cardiac surgery in the 1940s and 1950s led to accurate characterizations of etiologically defined conditions and paved the way for "discovery" of and interest in HCM. Evans (1) described several families with idiopathic hypertrophy and cardiac arrhythmias and grouped them as "familial cardiomegaly." In 1958, Teare (2), a forensic pathologist, reported autopsy findings on the young who died suddenly without known underlying disease. He was struck by the massive degree as well as the asymmetric nature of hypertrophy, and designated the condition as asymmetric hypertrophy of the heart. Brock (3), who had earlier done some pioneering work in functional infundibular pulmonary stenosis, described acquired subaortic stenosis in a patient with a previous history of hypertension. Goodwin and associates (4) in Britain observed clinical patterns in patients with idiopathic cardiac hypertrophy and subaortic obstruction, and used the term hypertrophic obstructive cardiomyopathy (HOCM). Almost in parallel, Braunwald and associates (5) in the United States embarked upon hemodynamic characterization of this disorder and called it idiopathic hypertrophic subaortic stenosis (IHSS). Also, simultaneously, Wigle and coworkers (6) in Canada further clarified understanding of clinical and hemodynamic parameters and used the term muscular subaortic stenosis (MSS). A keen interest in this disorder was rapidly generated and several important contributions were made by a number of investigatiors from different countries. A major fascination with this disorder (HOCM, IHSS, MSS) lay with a dynamic and variable left-ventricular outflow obstruction that seemed to respond paradoxically to most physiologic and pharmacologic maneuvers (7–10). Rapid progress and heightened interest led to an extensive review of the subject at a CIBA foundation symposium in 1964 (11). Beta-adrenergic blocking agents were used for clinical trials (12–15) and the role of surgery was better defined over the succeeding years (16–18). There was a flurry of excitement and activity when the concept of pressure gradients without outflow obstruction was proposed by investigators at the Johns Hopkins Hospital (19), believing that the pressure gradients in this disorder represented cavity obliteration rather than a meaningful obstruction to outflow of blood. This challenge was short-lived, as the presence of functional outflow obstruction was vindicated by careful studies (20, 21). It nonetheless served a useful purpose in pointing out the potentials for gradients other than outflow obstruction as well as for artifactual entrapment of catheters in a heavily trabeculated irregular left-ventricular cavity. In the late 1960s a new development in the form of diagnostic use of echocardiography generated further interest (22–24). This noninvasive tool helped with correct diagnosis in clinically unsuspected cases and confirmation in others without a need for invasive hemodynamic studies. These developments were reviewed in 1970 at the second CIBA foundation symposium on the subject (25). Longitudinal studies on clinical course and progression of this disorder were reported (26–29). A yet more recent development has been the use of echocardiography to diagnose asymmetric septal hypertrophy (ASH) and an attempt to designate the disorder as ASH on that basis (30). This is viewed as excessive reliance on one echocardiographic sign in the diagnosis of a complex disorder (31).

PATHOLOGIC FEATURES

Ventricular hypertrophy in patients with HCM may be either concentric or asymmetric. The heart with concentric hypertrophy is indistinguishable from that associated with systemic hypertension or aortic-valve stenosis. There is a uniform increase in muscle mass and wall thickness with little or no increase in cavity size. The appearance of the far more common asymmetric hypertrophy is quite distinctive (2). The distribution of increased muscle thickness is nonuniform, with massive increase in the interventrical septum and the adjacent anterior wall of the heart. The lateral and posterior free-wall thickness is less striking. Almost all patients with the clinical syndrome of hypertrophic obstructive cardiomyopathy (HOCM or IHSS) who have come to autopsy have demonstrated this striking asymmetric hypertrophy of the heart. A recent study (30) reports that the posterior-wall thickness directly behind the posterior mitral leaflet is significantly different in patients without outflow obstruction, in contrast with those with obstruction. This region in the "nonobstructors" was normal or less than normal in thickness, although the free wall caudal to and just lateral or medial to the posterior leaflet was thickened. In contrast, thickness of the free wall in this location in the "obstructors" was markedly increased. This is considered to represent myocardial hypertrophy in obstructors, in response to pressure overload from outflow obstruction. These pathologic observations are used along with and in support of echocardiographic evidence of asymmetric septal hypertrophy (ASH) in patients with hypertrophic cardiomyopathies. These reported differences remain to be confirmed by other investigators.

HISTOLOGY Teare (2) described the microscopic features as consisting of "a bizarre and disorderly arrangement of bundles of muscle fibers running in diverse directions, separated by connective tissue and clefts." He predicted with considerable foresight that "the connective tissue breaking up and interrupting muscle bundles, gives the impression of inefficiency in muscular contraction" of the massively hypertrophic septum and anterior wall. A recent study observed that the bizarre disarray of muscle fibers is localized to the i.v. septum in outflow obstructions, while it is more diffusely distributed in nonobstructors (32).

On the basis of gross pathology as well as microscopic features, the present NIH group of investigators (30, 32) has proposed that myopathic disease process is more extensive and diffuse in the nonobstructive form of HOCM than in the obstructive form. This conclusion is open to question, since the obstruction in this disease is characteristically variable and dynamic and those with outflow gradients up to 30 mm Hg were arbitrarily assigned in the nonobstructed group. Another major shortcoming of their data collection consisted of obtaining a necessarily smaller muscle sample for histology in a group of obstructed patients at surgery, whereas a large number of tissue blocks were examined in the autopsy group in which the nonobstructed patients predominated. Since the characteristic disarray of myocardial bundles is not diagnostic for this condition and since frequency of its detection is to some extent related to the number of tissue blocks examined, it appears that the very design of the studies favored finding a more diffuse involvement in the nonobstructed group of patients. Although these pathologic data may be viewed with considerable interest, the conclusions reached can only be considered tentative until further confirmatory studies are reported.

James & Marshall (33) examined hearts of 22 subjects who died suddenly and demonstrated asymmetric hypertrophy. Several histologic abnormalities were seen, none of which were universally present. Deep clefts in the septum were present in 7 hearts, the small coronary arteries were abnormally narrowed in 10, and the sinus node was sclerosed in 12. There were some abnormalities in all parts of the conduction system. A number of factors in pathogenesis of asymmetric hypertrophy were proposed, and include narrowed small coronary arteries, abnormal sequence of ventricular activation, and mechanical deficiency caused by deep clefts.

PATHOPHYSIOLOGY

Experimental and clinical studies on the genesis of cardiac hypertrophy have implicated a role of sustained increase in stroke work, of contractility, or of myocardial injury (ischemic or otherwise). The asymmetric hypertrophy of HCM may be viewed as a genetically determined distribution of increased muscle mass to any or all of the above factors. When the disease is clinically recognized, however, there is generally no obvious inciting factor. A number of studies have examined the hemodynamic consequences of bizarre generalized asymmetric hypertrophy.

Systolic Consequences

LV OUTFLOW OBSTRUCTION A profound systolic consequence of asymmetric hypertrophy is the development of a dynamic obstruction to left ventricular outflow in a majority of patients with HCM (hence the term HOCM). It differs strikingly from all forms of fixed obstruction in that it does not become manifest until after the early systolic ejection is under way. The severity of obstruction often varies from one beat to the next and from one phase of respiration to another (10). It is accentuated by factors associated with reduction in either preload of afterload, and with increase in contractility or heart rate. Owing to a variety of influences that tend to alter its severity, a patient may have a high gradient across left-ventricular outflow tract at one cardiac catheterization study and little or none at all at another time (34). Earlier selective angiographic studies (35) and more recent echocardiographic studies (22–24) have elucidated the mechanism of this obstruction as resulting from systolic anterior motion (SAM) of the anterior mitral leaflet with systolic reopening of the leaflet into the narrow outflow space. It is further demonstrated that the interventricular septum is inert or adynamic, possibly from bizarre disarray of muscle fibers with loss of efficient contraction. It is not clear as to why the anterior mitral leaflet reopens in midsystole, although the prevailing thought favors venturi effect resulting from a high-velocity outflow in early ejection. Another hypothesis is that alteration in geometry of the papillary muscle and insertion of mitral leaflet may result in an abnormally directed pull to the valve. Since a similar mitral-opening movement has been observed in other conditions with narrow outflow space or with high ejection velocity, the venturi effect in HOCM probably opens the valve, obstructs left ventricular outflow, and permits mitral regurgitation. As the forward flow is reduced, it permits the valve to return to a closed position in endsystole. Thus, the venturi effect necessary to accentuate SAM will produce or accentuate the severity of obstruction. A decrease in preload and afterload, as well as an increase in heart rate, may narrow the outflow space with reduction in distending pressure, whereas increased contractility and heart rate, and possibly reduced afterload, may permit a more rapid early-ejection velocity. Echocardiography has demonstrated increased and prolonged SAM of anterior mitral leaflet with amyl-nitrite inhalation, isoproterenol infusion, the Valsalva maneuver, and the postectopic beat. On the other hand, vasopressor agents (e.g. methoxamine, angiotensin) have produced a suppressed or absent SAM with amelioration of outflow obstruction. Similarly, when sustained afterload is present in a patient with asymmetric HCM in the form of fixed aortic valvar or supravalvar obstruction (including aortic coarctation) or systemic hypertension, SAM with associated dynamic outflow obstruction may be suppressed. Relief of distal obstruction may unmask the dynamic obstruction of HOCM.

Coexistence of mitral regurgitation was observed in earlier studies, with a prevalence estimated at about 50% of patients by ventriculography. A more consistent association of outflow obstruction and mitral regurgitation has been demonstrated by careful indicator-dilution studies (36). In most patients, provocative maneuvers that accentuate outflow obstruction will increase mitral regurgitation and vice versa. There are, however, a few patients (perhaps less than 10%) in whom mitral regurgitation is independent of obstruction and persists despite relief of outflow gradient. This may be related to actual deformity or damage to the mitral leaflets. Aside from these few instances, there is a close correlation between the presence and severity of dynamic-outflow obstruction and severity of regurgitation across the valve.

RIGHT-VENTRICULAR OUTFLOW OBSTRUCTION Although small pressure gradients across the right-ventricular outflow tract have been frequently reported, these are inconsequential in most instances and possibly even artifactual. Hemodynamically severe obstruction (gradient > 50 mm Hg) in the right ventricle may occasionally coexist with dynamic left-ventricular outflow obstruction in the same patient. Rarely, severe pulmonary infundibular stenosis may be isolated without any evidence of resting or provocable obstruction to left-ventricular outflow. Since the right-ventricular infundibulum is circumferentially bound by muscle, excessive contraction of the hypertrophied muscle results in outflow obstruction. Conditions that increase contractility tend to accentuate the severity of obstruction. Unlike the left-sided outflow obstruction, the tricuspid valve does not appear to play any role in right-side obstruction. No structural or functional defects of the tricuspid valve have been observed.

VENTRICULAR "PUMP" FUNCTION Systolic functional abnormalities of either ventricle are generally absent. Cardiac output is normal or even high at rest, and most patients respond appropriately to mild and moderate exercise. Peak rate of ventricular ejection is not diminished and may even be increased. The duration of systole is prolonged only in the presence of severe outflow obstruction, largely by prolongation of ejection time. Thus, either with or without outflow obstruction, total systolic pump function is relatively well maintained. Eventual deterioration in ventricular function with decrease in cardiac output and rise in filling pressures may take place during the end stage; however, this appears to be an uncommon development in the natural course of the disease.

Diastolic Consequences

Several studies have examined the diastolic function of the hypertrophied left ventricle. Elevated end-diastolic pressure associated with a small volume indicates poor compliance. Absence of a significant correlation between severity of outflow obstruction and the magnitude of end-diastolic pressure (34) is not surprising, since the bizarre hypertrophy in HOCM is independent of obstruction. Pressure-volume loops derived from frame-by-frame analysis of cineventriculograms are probably not reliable, since the grossly deformed and hypertrophied ventricular cavity is a far cry from the ellipse assumed in these derivations. A simple but effective approach has been to analyze left-atrial pressure pulse in diastole. The rate of diastolic pressure after mitral-valve opening from V to Y is sharply reduced, indicating poor atrial-emptying rate (or poor ventricular-filling rate) during passive diastolic filling period (37). Subsequent atrial systole is pronounced, with a prominent *a* wave, which is well-transmitted into the left ventricle. Visual frame-by-frame cineangiographic analysis with injection in the left atrium confirms the paucity of ventricular filling prior to atrial systole. Similar observations with right-atrial pressure pulse were made from the bedside evaluation of jugular-venous pulse. Recent emphasis on diastolic resistance to filling in HOCM has been placed by Goodwin and associates (38, 39). An observation of particular interest is the abnormal response of end-diastolic pressure to β-adrenergic stimulation. Isoproterenol infusion often raises end-diastolic pressure independent of its effects on outflow obstruction. Whether the rise in end-diastolic pressure is associated with a change in end-diastolic volume has not been clarified. Similarly, there are some data to suggest that β-adrenergic blockade lowers end-diastolic pressure with an increase in volume (40). Confirmation of these observations by other methods requires further studies.

Despite an increased resistance to filling, actual obstruction to inflow with a diastolic-pressure gradient between left atrium and ventricle is rare. Such instances may closely simulate mitral-valve stenosis.

CLINICAL RECOGNITION

Despite adequate descriptions of "typical" features of HCM with LV outflow obstruction (HOCM, IHSS), it continues to pose a diagnostic challenge at the bedside. Its presentation is varied, and clinical features are often confusing. The advent of echocardiography has made it a frequent "surprise diagnosis" in cardiology practice. Patients present with symptoms common to all forms of LV outflow obstruction i.e. dyspnea, angina, syncope, dizziness, and palpitations. Prolonged angina with a loud systolic murmur is often mistaken for manifestations of coronary artery disease. Younger patients free of symptoms are frequently referred for evaluation of a murmur. Features at the bedside that point to the diagnosis include (*a*) LV hypertrophy, (*b*) signs of dynamic LV outflow obstruction and mitral regurgitation, (*c*) signs of RV outflow obstruction, and, rarely, (*d*) signs of LV inflow obstruction.

LV Hypertrophy

Palpation of the left-ventricular apex in supine and/or left-lateral decubitus yields valuable information. Powerful systolic thrust of the left ventricle betrays an increase in muscle mass, and, although less frequent, the characteristic bifid apex in systole is virtually diagnostic. A prominent atrial-contraction wave imparts a strong presystolic impulse with a palpable *a* wave to the apex beat. Owing to a loss of compliance in HCM, the *a* wave can attain marked prominence. A trifid impulse composed of a prominent *a* peak and bifid systolic peaks are rarely appreciated at bedside, although readily recordable on apexcardiogram. An atrial sound (S4) is present in virtually all patients in sinus rhythm.

Signs of LV Outflow Obstruction

A jerky arterial pulse with sharp upstroke is typical, although not diagnostic. Occasionally a bifid pulse may be felt, especially in the carotid artery. Bifid arterial pulse in association with a normal pulse pressure is highly characteristic of HOCM. The initial peak results from an early unobstructed rapid ejection, and the later peak, presumably from relief of dynamic outflow obstruction as SAM of the mitral leaflet returns to its closed position. Pulse contour in HOCM is influenced by the presence and severity of outflow obstruction. In the absence of resting obstruction, arterial pulse is essentially normal.

A systolic murmur of variable intensity is present along the left sternal border and apex. It is poorly transmitted to the aortic area and neck vessels. It is medium- and high-pitched, with its onset after the first heart sound. The murmur resembles a long ejection murmur along the left sternal border and attains a regurgitant quality (high-pitched, blowing) out toward the apex. The second heart sound is clearly audible and both components are well-preserved. Reverse splitting with a delayed aortic component is diagnostic for a high-outflow gradient in absence of conduction delay. High-pitched ejection click is rare; however, a prominent lower-pitched sound (pseudoejection sound) often initiates the murmur. This sound cannot generally be distinguished from the onset of the murmur, but is readily recordable by phonocardiogram in approximately 50% of patients. A recent study (41) has shown that this sound bears a close temporal association with sudden halting of SAM of the anterior mitral leaflet as it strikes the interventricular septum. The pseudoejection sound is thought to result from a sudden deceleration of forward flow in the outflow tract with the onset of obstruction.

The signs of outflow obstruction, including intensity of the systolic murmur, are accentuated by maneuvers that augment the severity of obstruction. Since the dynamic-outflow obstruction is variable, so is the intensity of the murmur. Bedside evaluation for HOCM must therefore include response to the Valsalva maneuver and amyl nitrite inhalation.

The blowing apical murmur of mitral regurgitation varies in most patients with the dynamic outflow obstruction. As mentioned previously, accentuation of obstruction increases the intensity of the apical murmur and vice versa. In a few patients, associated severe mitral-valve regurgitation independent of outflow obstruction may be present. This can be determined by administration of a vasoconstricting agent (e.g. methoxamine or angiotensin), which while relieving outflow obstruction, does not diminish the murmur intensity. A surgical approach in these patients may require surgery on the mitral valve.

Signs of RV Outflow Obstruction

Ejection murmur of infundibular pulmonic stenosis (PS) is similar to that noted in congenital infundibular pulmonary stenosis. The murmur is most prominent at the lower left sternal edge. Ejection sound is absent and pulmonary-closure sound is delayed. The murmur intensity is accentuated by positive inotropic stimuli, but behaves differently in response to the Valsalva maneuver. It is softer during peak Valsalva effort, and intensifies following its release as the flow is increased. When infundibular obstruction accompanies LV outflow obstruction, the clinical signs of the latter dominate. However, predominantly isolated RV outflow obstruction may be difficult to differentiate from congenital infundibular PS, until evidence for unexplained left-ventricular hypertrophy is sought.

Signs of LV Inflow Obstruction

Whereas a prominent atrial sound (S4) is a constant feature of noncompliant hypertrophied left ventricle, a mitral diastolic murmur simulating mitral stenosis may occasionally result in diagnostic difficulty. Absence of opening snap and presence of unexplained left-ventricular hypertrophy should point to a correct diagnosis.

Without Obstruction

Patients with essentially similar symptoms and underlying pathology may show no signs of obstruction even on provocation. Presence of unexplained left-ventricular hypertrophy without any of the features of obstruction suggests the diagnosis of hypertrophic nonobstructive cardiomyopathy. These patients may have underlying concentric or asymmetric hypertrophy. When the latter is present, other family members of the patient may have typical features of HOCM.

DIFFERENTIAL DIAGNOSIS The common conditions misdiagnosed for HOCM include mitral regurgitation and coronary artery disease. Differentiation from aortic-valve stenosis offers little difficulty, although the murmur may simulate discrete subaortic stenosis. The latter is an example of fixed obstruction and behaves quite differently from the dynamic obstruction of HOCM to the provocative maneuvers. Differentiation from coronary artery disease is important, since the wrong diagnosis may lead to the wrong therapy, with potentially disastrous outcome. A HOCM patient with prolonged ischemic pain and hypotension and pronounced S-T changes on electrocardiogram may be mistaken for subendocardial infarction and pump failure with early cardiogenic shock. Use of positive inotropic agents may adversely influence the outcome by further accentuation of LV outflow obstruction. The correct diagnosis can indeed be made if the possibility is entertained, especially with the availability of echocardiography.

LABORATORY AIDS TO DIAGNOSIS OF HCM

Electrocardiogram (ECG)

A routine 12-lead ECG generally shows signs of left-ventricular hypertrophy with increased voltage and/or S-T-T wave changes. Despite a massive increase in muscle mass, it is surprising that LVH may be conspicuously absent. Thus a nearly normal ECG does not exclude HCM. Occasionally, increased electrical forces from the massively hypertrophied i.v. septum may produce prominent Q waves simulating infarction. These striking changes are present in only 10% of patients. A vectorcardiographic study (42) comparing chronic pressure- and volume-load hypertrophy with asymmetric hypertrophy failed to reveal any clear-cut differences. As a group, patients with volume-overload LVH tended to have increased "septal" rightward forces.

Other electrocardiographic features include a short P-R interval, WPW syndrome, left anterior hemiblock, and bundle-branch block. Atrial and ventricular arrhythmias may be present. The former may be episodic or permanent, and include atrial tachycardia or fibrillation.

Chest X Ray

Examination with PA and lateral chest X-ray films is generally unhelpful. Evidence of left-ventricular enlargement may be subtle, since the cavity size is not increased. Left-atrial size is either normal or only slightly increased, except in a stage of advanced decompensation. Pulmonary venous engorgement may be seen, but frank pulmonary edema and signs of pulmonary arterial hypertension are rare.

Echocardiography

This newer noninvasive technique has provided an important diagnostic aid in HCM (22–24). This technique permits a ready evaluation of the thickness of the interventricular septum and LV posterior wall (LVPW); their movements with systole; end-diastolic and end-systolic dimension of the LV cavity along its minor axis; LV outflow size, defined as space between anterior mitral leaflet and interventricular septum; and functional aspects of mitral and aortic valve motion (Figure 1). It also permits differentiation of concentric HCM from the asymmetric variety. In the former, the IVS/LVPW ratio is close to unity, while in the latter the ratio exceeds 1.3 (30). When the ratio of 1.3 is used, some degree of overlap may be observed; however, a ratio greater than 1.5 invariably denotes asymmetry.

In a substantial number of patients with documented asymmetric HCM, echocardiographic demonstration of ASH may not be technically feasible. Patients with either the concentric or the asymmetric variety of HCM have a normal or small LV cavity. The LV outflow size is narrower in the asymmetric HCM compared to the concentric type (normal=20–35 mm).

The presence of dynamic LV outflow obstruction is diagnosed by analyzing systolic motion of the mitral valve. Abnormal systolic anterior motion (SAM) of the anterior mitral leaflet with its apposition against i.v. septum localizes the outflow obstruction in HOCM.

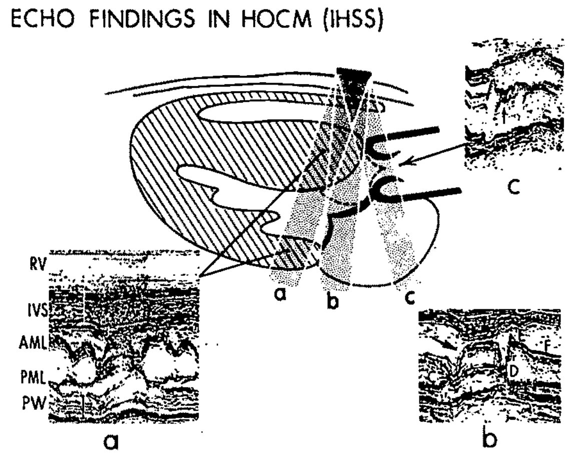


Figure 1 A schematic to illustrate by echocardiography the anatomic and functional changes in HOCM (*a*) When echo beam is directed as in "a" right ventricle (RV), i.v. septum (IVS), two leaflets of mitral valve (AML, PML), and posterior LV wall (PW) are recorded and serve to show asymmetric hypertrophy. The IVS/PW ratio in this example is 1.6. (*b*) With the echo beam directed as in "b," SAM (*arrow in the echo record*) of the anterior mitral leaflet can be seen as the opening movement (*dotted line in the schematic*) in systole with encroachment of the leaflet into LV outflow space. (*c*) Further angulation of the beam toward aortic root as in "c" shows midsystolic closure of the aortic valve cusps (*dotted line in the schematic and arrow in the echo record*) with reopening in late systole.

The SAM begins sometime after completion of early ejection, and is terminated in endsystole prior to the second heart sound. The dynamic nature of obstruction may be interpreted from variations in the extent of SAM with different maneuvers designed to alter obstruction. Patients without resting obstruction have small and incomplete SAM, while those with high resting gradients tend to have complete SAM consistently noted from one beat to the next. As a result of the dynamic midsystolic obstruction to outflow, the aortic-valve cusps may show premature closure with late systolic reopening. Diastolic movement of the mitral valve is impaired with a flat EF slope.

To summarize, a combination of narrow LV outflow space, thickened interventricular septum, and the typical SAM of the anterior mitral leaflet is virtually diagnostic for HOCM. When the interventricular septal-to-posterior wall ratio exceeds 1.5, asymmetric hypertrophy can be confidently diagnosed. Patients with HOCM, with rare exceptions, demonstrate asymmetric septal hypertrophy (ASH), although the latter is not specific for HOCM. Additional findings include midsystolic preclosure of one or more aortic-valve cusps, an inert i.v. septum with diminished systolic motion, and reduced diastolic EF slope of the anterior mitral leaflet. When SAM is absent at rest and following provocative maneuvers (i.e. Valsalva maneuver, amyl nitrite inhalation), it may be inferred that HCM, if present, is of the nonobstructive type. When multiple criteria are observed, a diagnosis can be made from the echocardiographic examination alone. However, in the absence of clear-cut evidence to support the diagnosis, further diagnostic evaluation may be undertaken.

Cardiac Catheterization—Angiography

Up until the advent of echocardiography, the final confirmation of a clinical suspicion rested with hemodynamic studies and selective angiocardiography. As discussed previously, a key diagnostic feature is the demonstration of dynamic LV outflow obstruction. However, special care must be taken to avoid the recording of an artifactual gradient from entrapment of the catheter. This can be recognized by careful attention to pressure-pulse contour. Any unexplained distortion of pressure pulse should be discarded. Occasionally, cavity obliteration may account for a high gradient between LV apex and arterial pressures without reflecting a true outflow obstruction. In order to obviate this, a transseptal catheter without side holes placed just distal to the mitral valve in the LV inflow should be used. Although this is a preferred approach in measurement of pressure gradients, transseptal catheterization is not without hazard. Most laboratories employ retrograde catheterization; however, it is important to avoid use of catheters with multiple side holes.

Selective left-ventricular cineangiography provides some characteristic features. Ventricular geometry is altered, with the cavity assuming a sausage shape in the RAO projection. Ejection fraction is normal or even increased. End-systolic volume is low, with areas of cavity becoming obliterated. A steep LAO projection may provide a view of the anterior mitral leaflet as it reopens in systole into the left-ventricular outflow space. Papillary muscles are prominent in the distorted LV cavity, with a thickened and bulging i.v. septum. Simultaneous left- and right-ventricular angiograms have been reported (43) in a few patients to demonstrate a massively thickened i.v. septum, especially in its midportion. These angiographic silhouettes are highly characteristic and provide measures (unreliable due to geometric distortion) of volumes, muscle mass, and ejection fraction. Since echocardiography generally provides a reliable means of diagnosing HOCM, catheterization-angiographic studies are generally reserved in selected cases for diagnostic confirmation, but more often in preparation for surgery.

MANAGEMENT

Since the etiology of this bizarre disorder is poorly understood, management is largely palliative. The major objectives are to prevent accentuation of outflow obstruction, improve left-ventricular compliance, and prevent and treat complication (i.e. cardiac decompensation, bacterial endocarditis) and suppress arrhythmias. It is hoped that with appropriate treatment, sudden death may be prevented and life prolonged.

Medical Management

The common cardiac drugs employed for symptomatic relief of dyspnea or angina in other cardiac disorders are either contraindicated or must be used with caution. Thus, use of digitalis must be avoided except for rapid atrial fibrillation when β-adrenergic blockade is unsuccessful or poorly tolerated. Nitrates are contraindicated and are generally ineffective in relieving angina. Diuretics may be used with caution, so as not to lead to hypovolemia.

Beta-adrenergic blocking agents, notably propranolol, have been extensively used. Slowing of the heart rate is generally beneficial. Although amelioration in obstruction induced by exercise and β-adrenergic stimulation has been demonstrated (13), resting gradients are reduced in only a few instances with labile obstruction. Symptomatic improvement is often striking and may last several years, although increasing doses may be needed (14, 15). Some patients fail to experience sustained improvement. Recent observations on improved LV compliance with β-adrenergic blockade (40), if confirmed, may provide a rational explanation for symptomatic relief. These agents may also have beneficial antiarrhythmic effects. The dose of propranolol should gradually be increased and carefully monitored, especially in those without a major component of outflow obstruction.

Surgical Management

Considerable experience has been reported with transaortic ventriculomyotomy and/or myectomy with muscle resection. Although the rationale for surgical relief of obstruction is poorly understood, several studies have documented postoperative relief in outflow obstruction both at cardiac catheterization and by echocardiography showing reduction or abolition of SAM of the anterior mitral leaflet. Symptomatic improvement is often dramatic, with reduction or abolition of systolic murmur and other features of LV outflow obstruction. A more recent surgical approach of mitral-valve replacement to relieve outflow obstruction is usually not indicated, except in a handful of patients with severe independent mitral regurgitation, since the standard ventriculomyectomy restores normal mitral leaflet function.

Surgical relief of RV outflow obstruction can be successfully carried out, as in infundibular pulmonary stenosis.

Other Measures

Prophylaxis for bacterial endocarditis should be instituted, since thickened anterior mitral valve or endocardial thickening in upper i.v. septum may serve as the locus of infection. Symptomatic antiarrhythmic drugs may be employed when indicated. Future studies are expected to better define their indications and role. Use of anticoagulants should be considered in the presence of atrial fibrillation, especially if systemic embolism is observed.

ETIOLOGY

Despite considerable progress in the understanding of clinical hemodynamic, pathologic, and functional aspects of HCM, underlying etiology and pathogenesis of development of the disease are shrouded in ignorance. The asymmetric type of HCM is almost certainly a genetically transmitted disorder, and the occasional sporadic cases are attributed to the failure to recognize subtle forms of the disease. A recent study on the genetic aspect, using an echocardiographic marker of asymmetric septal hypertrophy (ASH), suggested an autosomal-dominant transmission of the disorder (44). It remains to be seen whether the accuracy of echocardiographic diagnosis of ASH in otherwise healthy family members of patients proves reliable with the subsequent development of other manifestations of the disease.

Goodwin (45) has postulated a disorder of catecholamine function, possibly an abnormal response of myocardium to normal catecholamine actions, as a pathogenetic mechanism. Clinical association between pheochromocytoma, neurofibromatosis, and lentiginosis with HOCM may suggest a genetic disorder of neural-crest tissue. Experimental production of massive hypertrophy with obstruction following a small sustained infusion of norepinephrine may lend support to this postulate. Frequent clinical symptoms of autonomic dysfunction (i.e. excessive sweating) may also be noted.

It is possible that the manifest disorder represents a genetically determined response of hypertrophy to circulatory overload. This abnormal response to stimuli such as hypertension, athletic activities, and increased sympathetic or catecholamine action may be in the form of asymmetric hypertrophy with disarray of muscle fibers. Once the process is initiated, these abnormally arranged muscle fibers result in systolic and diastolic malfunction, which in turn may provide a stimulus for further hypertrophy.

NATURAL HISTORY

Longitudinal studies on the course of this disease are limited in the length of follow-up, since the medical community has become aware of this entity for less than two decades (26–29). The advent of echocardiography over the past seven years has permitted early detection and confirmation of HCM. Future long-term studies with additional anatomic and functional observations by this technique are expected to provide useful data on its natural course.

A recent multicenter study (29) of 190 patients with documented HOCM, having LV outflow gradient either at rest or on provocation, examined the progression of the disease with an average follow-up of 5.2 years. This was largely a retrospective study with participation by the Toronto General Hospital, University of Toronto, Royal Postgraduate School of Medicine and Hammersmith Hospital in London, University of Minnesota Medical Center, and University of Rochester School of Medicine and Strong Memorial Hospital. The major conclusions from this study were as follows:

1. Older patients tended to be more symptomatic (NYHA class III or IV), possibly indicating a progressive nature of the disease with time.
2. Clinical symptoms did not correlate with the severity of outflow obstruction or the murmur intensity, although there was no class I patient with a resting gradient in excess of 100 mm Hg.
3. After an average follow-up period of 5.2 years, 83% of the 141 surviving patients were either clinically improved or stable.

4. Sudden death was the most common single cause of demise—18% in the propranolol-treated group, 7% in the surgically treated (ventriculomyotomy) group, and 16% in the group receiving no treatment. A suggestion that successful surgery may protect against sudden death has not been borne out by statistical data.
5. There were no recognizable clinical predictors of sudden death, although it appeared to be uncommon after the age of 40 years.
6. No single symptom or combination of symptoms had a particularly ominous prognosis.
7. Progressive congestive heart failure with visceral congestion was rare, although this outcome may turn out to be more common with a longer follow-up of patients.
8. Twelve percent of patients had atrial arrhythmias, and established atrial fibrillation was observed in nearly half of these.
9. Overall mortality from complications of the disease was 18% over a follow-up period of 5.2 years, giving an attrition rate of approximately 3.4% per year.

The grim outlook on the basis of the frequency of sudden death has been supported by other studies. A more recent study in children gave an attrition rate of 4% per year (46). The disorder cannot be considered benign, and the danger of sudden death looms in the background despite clinical stability.

It must be pointed out that the above natural-history studies do not apply to patients with HCM without obstruction, although propensity to sudden death has also been noted in them.

The mechanism of sudden death has not been subjected to closer examination, although it is speculated to be of arrhythmic origin. Two mechanisms have been postulated: (*a*) rapid atrial arrhythmia may initiate ventricular tachycardia, or (*b*) primary ventricular arrhythmia rapidly degenerates into lethal events. A systematic quantitation of arrythmia by ambulatory monitoring is urgently indicated.

FUTURE STUDIES

A number of gaps remain in our understanding, with ample opportunity for clinical studies. A few of the questions to be examined are (*a*) is there a relationship between ambulatory arrhythmia and sudden death?; (*b*) will pharmacologic suppression of arrhythmias protect against the risk of sudden death?; (*c*) what is the role of LV outflow obstruction to sudden death?; (*d*) does successful ventriculomyotomy offer any protection against sudden death?; (*e*) what is the role of altered ventricular compliance to the clinical outcome?; and (*f*) can compliance be altered by pharmacologic agents?

Finally, it is well to remember that despite recent knowledge resulting from various clinical and functional expressions of the disease, it still represents an example of primary myocardial disorder diffusely involving myocardial cells, and the subsequent outcome of the disease may only be influenced by understanding the basic mechanisms that initiate and propagate it.

Literature Cited

1. Evans, W. 1949. Familial cardiomegaly. *Br. Heart J.* 11:68
2. Teare, R. D. 1958. Asymmetrical hypertrophy of the heart in young adults. *Br. Heart J.* 20:21
3. Brock, R. C. 1957. Functional obstruction of the left ventricle. *Guy's Hosp. Rep.* 106:221
4. Goodwin, J. F., Hollman, A., Cleland, W. P., Teare, R. D. 1960. Obstructive cardiomyopathy simulating aortic stenosis. *Br. Heart J.* 22:403
5. Braunwald, E., Morrow, A. G., Cornell, W. P., Aygen, M. M., Hillbish, T. F. 1960. Idiopathic hypertrophic subaortic stenosis: Hemodynamic and angiographic manifestations. *Am. J. Med.* 29:924
6. Wigle, E. D., Heimbecker, R. O., Gunton, R. W. 1962. Idiopathic ventricular septal hypertrophy causing muscular subaortic stenosis. *Circulation* 26:1153
7. Brockenbrough, E. C., Braunwald, E., Morrow, A. G. 1961. A hemodynamic technic for the detection of hypertrophic subaortic stenosis. *Circulation* 23:189
8. Braunwald, E., Ebert, P. A. 1962. Hemodynamic alterations in idiopathic hypertrophic subaortic stenosis induced by sympathomimetic drugs. *Am. J. Cardiol.* 10:489
9. Goodwin, J. F., Shah, P. M., Oakley, C. M., Cohen, J., Yipintsoi, T., Pocock, W. 1964. The clinical pharmacology of hypertrophic obstructive cardiomyopathy. See Ref. 11, p. 189
10. Shah, P. M., Yipintsoi, T., Amarasingham, R., Oakley, C. M. 1965. Effects of respiration on the hemodynamics of hypertrophic obstructive cardiomyopathy. *Am. J. Cardiol.* 15:793
11. Wolstenholme, G. E. W., O'Connor, M., eds. 1964. The cardiomyopathies: *CIBA Found. Symp.* London: Churchill
12. Cherian, G., Brockington, I. F., Shah, P. M., Oakley, C., Goodwin, J. F. 1966. Beta adrenergic blockade in hypertrophic obstructive cardiomyopathy. *Br. Med. J.* 1:895
13. Harrison, D. C., Braunwald, E., Glick, G., Mason, D. T., Chidsey, C. A., Ross, J. Jr. 1964. Effects of beta adrenergic blockage on the circulation with particular reference to observations in patients with hypertrophic subaortic stenosis. *Circulation* 29:84
14. Flamm, M. D., Harrison, D. C., Hancock, E. W. 1968. Muscular subaortic stenosis: Prevention of outflow obstruction with propranolol. *Circulation* 38:846
15. Adelman, A. G., Shah, P. M., Gramiak, R., Wigle, E. D. 1971. Long term propranolol therapy in muscular subaortic stenosis. *Br. Heart J.* 32:804
16. Morrow, A. G., Fogarty, T. J., Hannah, H. III, Braunwald, E. 1968. Operative treatment in idiopathic hypertrophic subaortic stenosis. *Circulation* 37:589
17. Bentall, H. H., Cleland, W. P., Oakley, C., Shah, P. M., Steiner, R. E., Goodwin, J. F. 1965. Surgical treatment and post-operative hemodynamic studies in hypertrophic obstructive cardiomyopathy. *Br. Med. J.* 27:585
18. Bigelow, W. A., Trimble, A. S., Auger, P., Marquis, Y., Wigle, E. D. 1966. Ventriculomyotomy operation for muscular subaortic stenosis. *J. Thorac. Cardiovasc. Surg.* 52:514
19. Criley, J. M., Lewis, K. B., White, R. I. Jr., Ross, R. S. 1965. Pressure gradients without obstruction: A new concept of "hypertrophic subaortic stenosis." *Circulation* 32:881
20. Wigle, E. D., Marquis, Y., Auger, P. 1967. Muscular subaortic stenosis: Initial left ventricular outflow tract pressure in the assessment of intraventricular pressure differences in man. *Circulation* 35:1100
21. Ross, J. Jr., Braunwald, E., Gault, J. H., Mason, D. T., Morrow, A. G. 1966. Mechanism of the intraventricular pressure gradient in idiopathic hypertrophic subaortic stenosis. *Circulation* 34:558
22. Shah, P. M., Gramiak, R., Kramer, D. H. 1969. Ultrasound localization of left ventricular outflow obstruction in hypertrophic obstructive cardiomyopathy. *Circulation* 40:3
23. Popp, R. L., Harrison, D. C. 1969. Ultrasound in the diagnosis and evaluation of therapy of idiopathic hypertrophic subaortic stenosis. *Circulation* 40:905
24. Shah, P. M., Gramiak, R., Adelman, A. G., Wigle, E. D. 1971. Role of echocardiography in diagnostic and hemodynamic assessment of hypertrophic subaortic stenosis. *Circulation* 44:891
25. Wolstenholme, G. E. W., O'Connor, M., eds. 1971. Hypertrophic obstructive cardiomyopathy. *CIBA Found. Sym.* London: Churchill
26. Frank, S., Braunwald, E. 1968. Idiopathic hypertrophic subaortic stenosis:

Clinical analysis of 126 patients with emphasis on the natural history. *Circulation* 37:759

27. Parker, B. M. 1969. The course in idiopathic hypertrophic subaortic stenosis. *Ann. Intern. Med.* 70:903
28. Adelman, A. G., Wigle, E. D., Ranganathan, N., Webb, G. D., Kidd, B. S. L., Bigelow, W. G., Silver, M. D. 1972. The clinical course in muscular subaortic stenosis. *Ann. Intern. Med.* 77:515
29. Shah, P. M., Adelman, A. G., Wigle, E. D., Gobel, F. L., Burchell, H. B., Hardarson, T., Curiel, R., de la Clajada, C., Oakley, C. M., Goodwin, J. F. 1974. The natural (and unnatural) history of hypertrophic obstructive cardiomyopathy: A multicenter study. *Circ. Res. Suppl.* II: 34, 35, pp. 179–95
30. Henry, W. L., Clark, C. E., Roberts, W. C., Morrow, A. G., Epstein, S. E. 1974. Differences in distribution of myocardial abnormalities in patients with obstructive and non-obstructive asymmetric septal hypertrophy (ASH): Echocardiographic and gross anatomic findings. *Circulation* 50:447
31. Shah, P. M. 1975. I.H.S.S., H.O.C.M., M.S.S., A.S.H? (Editorial). *Circulation* 51:577
32. Maron, B. J., Ferrans, V. J., Henry, W. L., Clark, C. E., Redwood, D. R., Roberts, W. C., Morrow, A. G., Epstein, S. E. 1974. Differences in distribution of myocardial abnormalities in patients with obstructive and non-obstructive asymmetric septal hypertrophy (ASH): Light and electron microscopic findings. *Circulation* 50:436
33. James, T. N., Marshall, T. K. 1975. De subitaneis mortibus: 12. Asymmetrical hypertrophy of the heart. *Circulation* 51:1149
34. Braunwald, E., Lambrew, C. T., Rockoff, S. D., Ross, J. Jr., Morrow, A. G. 1964. Idiopathic hypertrophic subaortic stenosis. I. A description of the disease based upon an analysis of 64 patients. *Circulation* 30:3
35. Simon, A. L., Ross, J. Jr., Gault, J. H. 1967. Angiographic anatomy of the left ventricle and mitral valve in idiopathic hypertrophic subaortic stenosis. *Circulation* 36:852
36. Wigle, E. D., Adelman, A. G., Auger, P. 1969. Mitral regurgitation in muscular subaortic stenosis. *Am. J. Cardiol.* 24:698
37. Shah, P. M., Gramiak, R., Kramer, D., Yu, P. N. 1968. Determinants of atrial (S4) and ventricular (S3) gallop sounds in primary myocardial disease. *N. Engl. J. Med.* 278:753
38. Goodwin, J. F. 1974. Prospects and predictions for the cardiomyopathies. *Circulation* 50:210
39. Oakley, C. M. 1971. Hypertrophic obstructive cardiomyopathy—Patterns of progression. See Ref. 11, p. 9
40. Webb-Peploe, M. M., Croxson, R. S., Oakley, C. M., Goodwin, J. F. 1971. Cardioselective beta adrenergic blockade in hypertrophic obstructive cardiomyopathy. *Postgrad. Med. J. Suppl.* 47:93
41. Sze, K. C., Shah, P. M. 1976. Pseudoejection sound in hypertrophic subaortic stenosis: An echocardiographic correlative study. *Circulation.* 54:504–9
42. Brackbill, T. A., Shah, P. M. 1974. Vectorcardiographic comparison of left ventricular hypertrophy in idiopathic hypertrophic subaortic stenosis, aortic stenosis, and aortic regurgitation. *Am. Heart J.* 88:269
43. Redwood, D. R., Scherir, J. L., Epstein, S. E. 1974. Biventricular cineangiography in the evaluation of patients with asymmetric septal hypertrophy. *Circulation* 49:1116
44. Clark, C. E., Henry, W. L., Epstein, S. E. 1973. Familial prevalence and genetic transmission of idiopathic hypertrophic subaortic stenosis. *N. Engl. J. Med.* 289:709
45. Goodwin, J. F. 1975. ?IHSS. ?HOCM. ?ASH. A plea for unity (Editorial). *Am. Heart J.* 89:269
46. Maron, B. J., Henry, W. L., Clark, C. E., Redwood, D. R., Roberts, W. C., Epstein, S. E. 1976. Asymmetric septal hypertrophy in childhood. *Circulation* 53:9

Ann. Rev. Med. 1977. 28:251–69
Copyright © 1977 by Annual Reviews Inc. All rights reserved

CURRENT PROBLEMS IN MYCOSIS FUNGOIDES AND SÉZARY SYNDROME

❖7241

R. K. Winkelmann, M.D., Ph.D.
Departments of Dermatology and Anatomy, Mayo Medical School; and Mayo Clinic and Mayo Foundation, Rochester, Minnesota 55901

W. A. Caro, M.D.
Department of Clinical Dermatology, Northwestern University Medical School, Evanston, Illinois 60201

MYCOSIS FUNGOIDES

Since Alibert described mycosis fungoides in 1806, there has continued to be confusion regarding this disease. Despite recent reviews (1–3), much of the confusion persists and only in the last 2 years have cellular and immunologic studies provided additional significant data. We would like to list the following significant questions regarding mycosis fungoides (MF) and then try to provide current answers to them: (*a*) what is MF?, (*b*) what is the clinical variation of the disease?, (*c*) what is the cellular nature of MF?, (*d*) what is the histopathology of MF?, (*e*) what is the immunologic status of MF?, (*f*) what are the staging procedures for evaluation of MF?, and (*g*) what is adequate therapy for MF? A critical review of the data and the opinions derived from them will help set the stage for more progress in the management of MF.

Definition

It does not seem possible that physicians still define MF in disparate ways, yet from Symmers (4) to Reed & Cummings (5) it has repeatedly been stressed that MF does not exist, that it is at most a lymphoma involving the skin. Many recent authors, for example Clendenning (6) and Lutzner et al (7), state that MF is any lymphoma of the skin. The classic description of MF was enunciated once again by Samman (8, 9) and reflects the attitude of many dermatologists that MF is best thought of as a chronic inflammatory and potentially prelymphomatous cutaneous syndrome that presents in three stages: the parapsoriasis or dermatitis stage, the infiltration stage, and the tumor stage. In this view, only the tumor stage is considered lymphoma. Tan et al (10) state that

> . . . mycosis fungoides (MF) is a chronic granulomatous response to persistent unidentified antigen(s), upon which immune imbalance can develop resulting in "auto-immune" phenomena and, in a few cases, the emergence of various lymphoreticular neoplasms.

Winkelmann (11) states that MF is a chronic benign epidermotropic lymphoproliferative reaction that may develop into true lymphoma. The pathologists, such as Symmers (4), who view MF from the autopsy table, recognize that patients dying of cutaneous lymphoma may belong to a heterogeneous group that defies rigid morphologic classification. At the National Cancer Institute (NCI) in the United States, severe tumor cases of MF are often available for study, so the features of lymphoma are always the most apparent and important. The clinical dermatologist recognizes the variable and slowly developing benign course and histology of the premycosis patient and continues this view to include the infiltration stage, for most such patients respond to simple therapy, have a good prognosis, and may even have spontaneous healing of skin lesions.

The injection of mice with allogeneic cells produces inflammation, lymphoproliferation, and finally, if the animal survives, lymphoma (12). This is identical to the stages of MF and its evolution to malignancy. The malignant stage in mice has been associated with virus, particularly in the NZB mice. Therefore, experimental immunologic and oncologic models for MF exist that involve the transition from inflammation, through immunoreactive proliferation, to lymphoma. In man, benign inflammatory cutaneous lymphocytic reticulosis shows the same features. Lymphocytoma, acrodermatitis chronica atrophicans, and Sézary syndrome are clinical variants of dermal lymphocytic reticulosis that may develop into lymphoma (Table 1). It is tempting to consider true MF only as the transitional infiltrative stage II Alibert disease, and to regard previous inflammation as premycosis and subsequent lymphomas as a definite change to tumor.

Clinical Manifestations and Course

The typical three stages of MF were not outlined by Alibert, but were clearly defined by Bluefarb (13), who rejected the erythrodermic form and the d'emblée form as usually representing lymphoma cutis, not MF. Problems of identification based on

Table 1 Cutaneous lymphoreticular proliferation

Benign	Lymphocytoma Acrodermatitis chronica atrophicans Parapsoriasis
Transitional	Mycosis fungoides Sézary syndrone Eosinophilic reticulosis Extramedullary plasmacytoma
Malignant	Lymphoma

clinical and histopathologic grounds exist, for in the series of Degos et al (14) 60% of patients had the erythrodermic form, whereas in that of Epstein et al (1) only 10% had this form. In our experience, erythrodermic MF does exist as a rare, generalized, confluent extension of infiltrative MF. However, we agree with Bluefarb (15) that most cases of "red man" represent some other dermatosis or dermatitis, a drug reaction, lymphoma, or Sézary syndrome. The 10% of cases of MF d'emblée can be considered as systemic lymphoma with cutaneous tumor involvement, for visceral lesions are found within months, if they are not evident when the patient is first observed. Although such rapid change is typical of lymphoma, it is not the classic MF. We agree with Bluefarb (15) that such cases should be accepted as representing "accelerated" MF only if the other stages are clinically and histologically visible on the skin at the same time and staging procedures are negative.

Typical MF has a protracted course, although when the tumor or ulcerative stage is reached, it is often only a matter of 2 years or less until death. Individual lymphomatous tumors may occasionally be observed and cured with intense X-ray therapy during stage I and stage II MF without affecting the course of the basic process or the prognosis. Samman (8) has emphasized that the skin lesions of premycosis cannot be accurately diagnosed by clinical and histologic criteria and that therefore a change occurs when typical stage II MF or stage III lymphoma develops.

For many authors, MF is always the same disease in all stages, whether it can be recognized and separated from parapsoriasis or other dermatoses or not. Similarly, Rappaport & Thomas (16) and Long & Mihm (17) believe that MF is still MF in the tumor lymphomatous stage. They emphasize the similarity of the histologic picture in the lymph nodes, spleen, and other visceral organs. These authors do not believe that MF develops into other forms of lymphoma. Allen (18), Gall (19), Montgomery (20), Samman (9), and many others believe that reticulum-cell sarcoma (histiocytic lymphoma), lymphosarcoma, Hodgkin's disease, and leukemia of several cell types may occur. What is most important is that all authors agree that lymphoma can develop in MF, whether at onset or later, and that when MF does manifest itself in tumors, ulcers, or lymphadenopathy, the prognosis is often limited to one or two years' survival.

The Cytology of Mycosis Fungoides

MF has been described as a polymorphous reticulosis of the skin with a mixed cellular infiltrate composed of lymphocytes, histiocytes, and occasional eosinophils, neutrophilic leukocytes, and giant cells. The infiltrate has been characterized by a large mononuclear cell with a hyperchromatic nucleus that is found in the dermis and epidermis. This cell has been termed the "mycosis" cell, but no further determination of its origin has been possible. Touch preparations of MF lesions usually reveal a mixture of lymphocytes, monocytes, and eosinophils (21).

Electron microscopy of MF lesions shows two cell populations: the lymphocytes and the larger reticular cells. Orfanos & Stüttgen (22) found large reticular cells with many organelles and inclusions. Bonneau & Cesarini (23) studied all three stages of MF and noted in the premycosis stage a cellular polymorphism and no malignant cells, in contrast to stage III, in which monomorphism of malignant reticular cells was the rule. In stage II these authors found the infiltrate to be a mixture of "MF" cells and histiocytes. Lutzner & Jordan (24) described the ultrastructure of the Sézary cell as a lymphocyte with a hyperconvoluted nucleus. Lutzner and associates (25) noted that the same cells were to be observed in MF, and these authors indicated that this cell might be the mycosis cell. It has generally been agreed that the lymphoid cells with hyperconvoluted or cerebriform nuclei are typical of, but not specific for, MF. Fisher et al (26) noted that the cells of MF were qualitatively similar in all stages and that there was no qualitative difference between MF and other dermatoses. Fisher et al (26) also stated that the MF cell is a reticulum cell and found many intranuclear bodies, as did Sandbank (27). Flaxman et al (28) noted cerebriform cells in the cellular infiltrates of lichen planus, solar keratoses, discoid lupus erythematosus, vasculitis, and acute guttate parapsoriasis. Thus, electron microscopy has not settled the origin of the MF cell. Some believe the MF cell to be the cerebriform lymphocyte and others regard it as a reticulum cell. Both cell types characterize the ultrastructural appearance of MF. Rosas-Uribe et al (29) indicated that masses of hyperconvoluted cells are indicative of MF.

Burg & Braun-Falco (30, 31), using cytologic and cytochemical techniques, also observed two cell types in the lesions of MF. In normal skin of MF patients, they found an increase in monocytes, histiocytes, and macrophages in the papillary dermis. Later, perivascular reticulum cells were observed. Vesper et al (32) studied MF and parapsoriasis by the skin-window technique and described an early benign monocyte population that appeared at the skin window. No antigen stimulus was used. These benign masses of monocytes appeared as early as 3 hours. Similar preparations of lymphoma nodules produced masses of malignant cells with heterochromatin and large nucleoli. Increased monocytopoiesis in the bone marrow and peripheral blood has been demonstrated by Meuret et al (33) with the tritiated thymidine technique. The promonocyte pool was enlarged and the labeling index of the promonocytes increased to four to five times normal values. An increase in immature and pyknotic monocytes and a new transitional monocyte were observed in the peripheral blood. These authors believe that their data support the concept that MF is associated with a systemic disorder of the monocyte-macrophage system. Seitz et al (34) demonstrated defective monocyte chemotaxis in MF. Burg & Braun-Falco (31) also found monocytoid cells in the blood of MF patients. Exfoliative cytology of MF has been studied by Goldschmidt & Thew (35), and although 2 of 18 patients had histiocytes, minimal or no inflammatory or infiltrative cells were observed. The sea-blue histiocyte has been observed in MF (36). Brehmer-Andersson & Brunk (37) used tape stripping for cytologic diagnosis of MF and found malignant mononuclear cells with heterochromatin and abnormal nucleoli. Several cells were large enough to be considered tumor giant cells of the Sternberg-Reed type.

Intraepidermal reticulosis has been described under the eponym of Woringer-Kolopp disease. Clinically, it is typically described as a benign, localized, scaling plaque, usually on the foot, although a generalized malignant form has been recognized. Microscopically, this entity has been compared to MF or intraepithelial reticulum-cell sarcoma (38). Merkel cells have been found in the epithelium of patients with this disease by Braun-Falco et al (39) and by Revuz et al (40). It has been proposed that this disease, which is microscopically identical to MF, is a "Merkeloma."

The study of cells from touch imprints of MF lesions stained by the Feulgen method and recorded photoelectrically indicates the presence of hyperdiploid cells. Van Vloten et al (41) believe that the presence of such cells indicates malignancy early in the course of the disease, and that this procedure can distinguish between nonspecific dermatitis simulating MF, and true MF in the early stage. Because Sézary cells have been shown to be hyperdiploid and present in MF lesions as well as other diseases, such as lichen planus, such cells may account for the findings of van Vloten et al (41). Cytogenetic studies by Erkman-Balis & Rappaport (42) on 11 patients with MF showed abnormal karyotypes in four skin biopsies, marker chromosomes, and a modal chromosome number ranging from 44 to 50. The chromosome changes remained constant in a given case for periods of up to 18 months. Similar findings were observed in cells from the lymph nodes and visceral organs. These findings could not be related to treatment by the authors, nor were they changed by treatment. Studies of this type must be related to the extent and stage of the disease, and to date they provide the best evidence of neoplastic cells in MF.

The cytologic picture in MF appears to be a polymorphous response composed principally of lymphocytes and larger reticular cells. Sheets of cerebriform lymphocytes appear to be characteristic of MF, as does the appearance of a monomorphous malignant reticular cell population. Early stages may show lymphocytes only, as in parapsoriasis, or inflammatory cells, as in dermatitis. Increased activity of the lymphocyte and monocyte populations has been measured by special studies. An unresolved and important question is whether the benign and malignant reticular and histiocytic cells observed in MF develop from T lymphocytes. All of the cellular reaction of MF could be a T-cell proliferation, at first benign and later malignant, or malignant from the onset despite the long course and benign histology.

Histopathology of Skin Lesions

The earliest, or eczematous, stage of MF continues to present difficulties in microscopic diagnosis, and a clinical picture strongly indicative of MF may antedate the appearance of diagnostic histopathologic features (W. A. Caro, D. Variakojis, unpublished data). These early lesions show a variable microscopic picture of chronic dermatitis, which may be psoriasiform, poikilodermatous, or nonspecific. The diagnosis of MF may be suggested in these early stages by the identification of atypical mononuclear cells with irregular nuclear configurations and frequent hyperchromasia. Because of the lack of specificity of the early histopathologic picture, there are no histopathologic criteria for determining the precise onset of MF.

In the infiltrative plaque stage, the skin biopsy usually reveals the diagnostic features of MF (43). The upper dermis shows a dense, often bandlike cellular infiltrate with discrete patches of cells lying at all levels of the dermis. The infiltrate may contain a multiplicity of cell types, but the basic cell is an abnormal-appearing mononuclear cell ranging in size from a medium to a large lymphocyte. The cytoplasm is usually inconspicuous, and the nucleus has an irregular configuration. Larger cells may have irregularly indented or hyperconvoluted nuclei. Some cells have markedly hyperchromatic nuclei (mycosis cells); others are paler and show a pattern of coarse and irregular clumped chromatin with interconnecting strands. Although the larger cells may appear frankly pleomorphic and neoplastic, smaller cells may have a more uniform and bland appearance (16).

The typical MF plaque also has a variable infiltrate of normal-appearing lymphocytes, histiocytes, plasma cells, and eosinophils (43).

The cells of MF have a special affinity for the epidermis, and this epidermotropism is an essential diagnostic criterion for cutaneous MF. The Pautrier microabscess can be identified in most biopsies of MF plaques and appears within the epidermis as a round clear space within which one finds the same types of cells as in the dermal infiltrate. Pautrier microabscesses vary in number, and a large number of sections may be required before one is found.

Although most investigators consider the Pautrier microabscess to be characteristic of MF (16, 43), occasional reports cite the association of Pautrier microabscesses with other conditions. Kim et al (44) and Petrozzi et al (45) identified Pautrier microabscesses in cases they diagnosed as reticulum-cell lymphoma, and Ackerman and associates (46) have found structures closely resembling Pautrier microabscesses in benign chronic dermatitis. Similar structures have also been identified in lymphomatoid papulosis (47–49).

The other form of characteristic epidermal involvement by MF consists of a more diffuse invasion of the epidermis by cells of the infiltrate. In this type of exocytosis, individual cells or small groups of cells are surrounded by clear spaces (16), and this change is usually easily distinguished from the spongiosis and exocytosis of dermatitis.

In the tumor stage of MF the infiltrate has become more massive and usually occupies the entire dermis, with frequent extension into the subcutaneous fat (43). The epidermis may undergo extensive infiltration and ulceration. The abnormal cells appear in much larger numbers, and the inflammatory cells may be totally suppressed. Some cases show considerable cellular pleomorphism with large numbers of bizarre-appearing tumor cells, and such cases may be difficult to differentiate from histiocytic lymphoma (16, 43). Other cases may closely resemble Hodgkin's disease with cells indistinguishable from Sternberg-Reed cells, and still other cases may suggest a form of lymphocytic lymphoma with diffuse proliferation of relatively uniform-appearing lymphoid cells.

Histopathology of Extracutaneous Involvement

Extracutaneous involvement usually occurs late in the course of MF, and the likelihood of such involvement increases as more extensive and more infiltrated skin lesions develop (1, 3). The reported incidence of extracutaneous involvement varies from as low as 15% (18) to more than 80% (1, 50). These wide variations reflect in part the variations of the disease and in part differing criteria used to identify MF. Prolonging life by preventing secondary infection of skin lesions may allow time for extracutaneous spread to develop (16).

Besides the skin, specific involvement occurs most commonly in lymph nodes, followed, in varying order of frequency, by lung, spleen, liver, and kidney (1, 16, 17, 50, 51). Bone-marrow involvement occurs somewhat less often, although virtually any organ may be affected.

Rappaport & Thomas (16) have listed histopathologic criteria common to all areas of involvement. The cellular proliferation consists of atypical mononuclear cells that are essentially the same as those in the skin. Some distinctive cells have deeply infolded nuclei, and others have large hyperchromatic nuclei that may or may not be indented (mycosis cells). The cellular infiltration of organs and tissues is usually less destructive than that in other malignant lymphomas, and the underlying architecture is often well preserved. The affinity of the cellular infiltrate for epithelial structures also occurs in other areas of epithelium, such as the renal tubules or bladder epithelium (16).

Rappaport & Thomas (16) believe that the cytology of the infiltrate, the distribution in extracutaneous tissues, and the cellular pattern within an organ are characteristic for MF and do not suggest other malignant lymphomas. Less classic extracutaneous involvement may be more difficult to differentiate from other malignant lymphomas with cellular patterns resembling Hodgkin's disease or histiocytic or lymphocytic lymphoma. Although many authors believe that these patterns represent transition of MF into other malignant lymphomas (51, 52), others regard them as variants of MF itself and deny that such transition takes place (16, 17).

Involvement of lymph nodes, the most common extracutaneous site, presents special problems (16). Involvement ranges from dermatopathic lymphadenitis without evidence of MF to nodes with total involvement. Nodes with focal involvement can also be identified as showing MF. Dermatopathic lymphadenitis may resemble MF nodal involvement, because the extensive proliferation of the larger histiocytes replaces normal nodal tissues in a manner similar to that in MF. The differentiation from lymphocytic lymphoma and histiocytic lymphoma may be difficult, but in MF one often finds areas of residual normal lymphoid tissue even in diffusely involved nodes (16, 53). Identification of the characteristic cells with their infolded nuclei also helps to distinguish MF from other malignant lymphomas (16). Small numbers of the cells may be present in dermatopathic lymphadenitis, but their presence in clusters or sheets is indicative of MF (29).

The diagnosis of dermatopathic lymphadenitis in MF carries with it a serious prognosis, even though MF cannot be demonstrated in the lymph nodes. Patients with dermatopathic lymphadenitis have diminished survival when compared with MF patients who do not have nodal involvement (2, 54). This finding suggests that dermatopathic lymphadenitis may serve as an indicator that extracutaneous spread has occurred, even though it cannot be identified at that time. Karyotype abnormalities have been demonstrated in lymph nodes with dermatopathic lymphadenitis in patients who have MF, and this further suggests that these nodes are already involved (42).

In studies of extracutaneous involvement, there has been a close correlation between lymph-node involvement and visceral involvement. In the series of Variakojis and associates (53), two of three patients with MF demonstrated in lymph nodes had splenic involvement. Rappaport & Thomas (16) reported that all 24 autopsied patients with lymph-node MF had visceral involvement, and 4 patients without demonstrable lymph-node involvement also had visceral involvement. These autopsy data suggest that the barrier to the spread of MF may lie within the skin and that lymph nodes are affected in a sequence similar to that in other extracutaneous sites (16). Thus the failure to demonstrate MF in lymph nodes gives no assurance that the patient lacks other extracutaneous involvement, and the finding of MF in lymph nodes strongly indicates other extracutaneous involvement.

Immunopathology

The patient with MF may be healthy for years, and during this time normal delayed-hypersensitivity skin tests may be found (55). The patients are readily sensitized to locally applied dinitrochlorobenzene (56) and, during treatment, to nitrogen mustard (57). Drug reactions and contact dermatitis occur readily in patients with MF and are often the cause of erythroderma. Direct immunofluorescence does not reveal cells with surface immunoglobulins, although Lai A Fat and associates (58) have demonstrated occasional cells that they believe synthesize immunoglobulins in the dermal infiltrate of MF. Occasionally, a band of immunoglobulins is deposited at the dermo-epidermal junction in an MF lesion (59). It is easy to extract masses of cells with the T-cell property of forming sheep RBC rosettes from lesions of MF. These cells can be stimulated by phytohemagglutinin (60, 61).

Clendenning et al (62) observed a few cerebriform lymphocytes in the blood of 20% of NCI patients with MF. In only 1 of 150 cases of MF have Sézary cells been observed at the Mayo Clinic. The peripheral lymphocytes from MF in stage I and stage II showed normal response on phytohemagglutinin stimulation (63). In the tumor state, an inhibition of DNA synthesis and blast transformation were found. Lutzner et al (7) found that the lymphocytes of MF patients usually respond normally to nonspecific mitogens, specific antigens, and mixed-lymphocyte culture. Only when delayed-hypersensitivity tests failed (stage III lymphoma) were there alterations in responses. Umbert et al (64) noted circulating macrophage migration inhibition factor in patients with stage II MF, presumably correlated with T-cell activity, but migration inhibition factor was not found in stage I or stage III MF. The number of circulating T lymphocytes in MF is usually normal or slightly depressed. There is often an increase in circulating B lymphocytes, particularly with IgM and IgD receptors, according to Carapeto et al (65). Tan et al (10) found normal B-cell values, in contrast to the study of Carapeto et al, wherein only two of seven patients had normal numbers of blood B lymphocytes. Confirming B-cell changes, Kawacki et al (66) found serum γ-globulin concentration to be higher in MF than in lymphoma. Significant increases in IgG and IgA were found in all cases. Blaylock et al (55) found elevated IgA in 8 of 12 patients with MF. Tan et al (10) found an elevation of IgE in all patients studied, although only two were atopic. Polyclonal gammopathy of the IgG-IgM type has been observed. Herrmann & Gartmann (67) reported finding cryoglobulins in two of eight cases of MF. Organ-specific antibodies were observed in 13 patients by Tan et al (10), and antinuclear antibody was found in 10 of these 13 patients.

Staging

Information derived from staging procedures in Hodgkin's disease and non-Hodgkin's lymphoma has clearly indicated that not all cutaneous lymphoma is stage IV (68). Localized primary cutaneous lymphoma with a serious prognosis and that with a favorable prognosis can and must be distinguished for proper therapy. Histologic classification of non-Hodgkin's lymphoma in general has revealed a relation between histologic type and prognosis (69). Kim and associates (44) could not find a difference in prognosis among cases of histiocytic, monocytic, or undifferentiated reticulum-cell sarcoma of the skin, but the numbers of cases were too small to be treated statistically. The modified Ann Arbor classification notes that extranodal involvement may be stage I–localized; stage II–localized in two or more sites on the same side of the diaphragm; stage III–localized on both sides of the diaphragm; and stage IV with diffuse involvement (70). Winkelmann (71) proposed a staging classification created specifically for skin lymphoma that recognizes stage I focal cutaneous lymphoma, stage II limited to contiguous areas, stage III generalized cutaneous involvement only, and stage IV cutaneous involvement with visceral lymphoma. Regional lymph-node involvement associated with focal skin involvement was suggested as a subset within that particular stage and indicated that no visceral nodes were present.

Staging examinations include complete history and physical and laboratory examination, including skin biopsy. Blood and marrow studies should include the hemoglobin concentration, the RBC, WBC, differential, and platelet counts, a special smear for Sézary or other atypical cells, and a bone marrow biopsy (50, 72). The bone marrow examinations have usually been negative. Lymphocytopenia often occurs at some time in the course of MF and may represent transition to lymphoma, for Fuks et al (54) found it to be a poor prognostic sign. Delayed hypersensitivity skin tests and sensitization with dinitrochlorobenzene are important procedures, for the depression of this response correlates with a poor prognosis.

X-ray examinations for staging include studies of the chest, gastrointestinal tract, and renal system. Liver and renal function studies and routine blood chemistries are required. The technetium liver and spleen scan is useful, but an enlarged liver or abnormal liver function often does not indicate microscopic involvement in either MF or Sézary syndrome. The enlarged spleen is an important finding, and it may contain microscopic MF, as in the study of Griem et al (72). The largest normal spleen was 155 g, and of five abnormal spleens at laparotomy the weight was more than 250 g in all. Three of these spleens had an abnormal scan. Whole-body and bone scans have not been helpful to date.

Palpable lymph nodes are most often found to be dermatopathic lymphadenopathy on microscopic examination. Lymph-node biopsy of palpable lymph nodes is necessary, and serial biopsy is not inappropriate to monitor progression of the disease. MF cells are found in lymph nodes, and their presence without destruction of lymph-node architecture and associated with additional cytologic features of malignancy must be further evaluated. Positive findings in regional lymph nodes indicate a poor prognosis; Fuks et al (54) noted that palpable nodes of any histologic type indicated a poorer prognosis. On this basis, it is recommended that every patient with MF have a biopsy of palpable lymph nodes or that a blind lymph-node biopsy of nonpalpable nodes be done.

Lymphangiography has been found to be helpful in assessing lymphadenopathy in Hodgkin's disease. This study provides useful help in assessing retroperitoneal lymph node involvement and should be done before laparotomy (54). Although both false-positive and false-negative results are observed, it has an established place in the staging of lymphoma. Escovitz et al (73) studied 44 patients who had biopsy-proved MF, and found 10 of 25 with no clinical evidence of lymph node disease who had significant lymphangiographic evidence of lymphoma. Serial lymphangiograms were negative, but at the Mayo Clinic, Winkelmann has noted the pseudolymphoma change produced by serial lymphangiography. Inferior venacavography can also demonstrate involved lymph nodes (72).

Laparotomy for staging MF has not been performed often, and only two reports on the same patient population exist (from the University of Chicago) (53, 72). Laparotomy is important for extranodal lymphoma, because 60% of such patients do have visceral disease, the pattern of the disease is not the same as that in nodal Hodgkin's disease, and decisions regarding therapy are often influenced by the specific findings. Then too, it is not unusual to find negative laparotomies in patients whose previous staging studies indicated visceral lymphoma. The very elderly, debilitated patients, patients with severe secondary infections, and patients with other histologic evidence of visceral disease should not be considered candidates for laparotomy. Variakojis et al (53) and then Griem et al (72) reported on 14 patients with various clinical presentations who underwent laparotomy. Clinical skin morphology did not correlate with laparotomy findings. MF was found in the lymph nodes in three, in the spleen in four, and in the liver in two patients. All patients had chemotherapy, irradiation, and considerable local therapy for 1–2 years before laparotomy, and five patients had tumors or generalized plaques. These investigators believe that patients with lymphadenopathy, splenomegaly, or abnormal spleen scans or lymphangiograms should be considered for staging laparotomy. The finding of convoluted cells in lymph nodes, spleen, and other organs can occur in two ways: (*a*) localized proliferation in the T-cell areas as in the skin and (*b*) generalized proliferation with tissue destruction and cellular anaplasia. These two qualities should be clearly distinguished in future studies and related to the stage of the disease and its outcome.

Treatment of Mycosis Fungoides

Levi & Wiernik (3) have reviewed the therapy of MF, and the three general modes of treatment include topical therapy, radiotherapy, and systemic chemotherapy. Topical therapy and radiotherapy are indicated when involvement is limited to the skin, and systemic chemotherapy is employed for more extensive disease.

In recent years, investigators have been adopting a more aggressive attitude toward the treatment of MF (3, 57, 74–76), based on the following assumptions:

1. MF is a malignant lymphoma and, along with other malignant lymphomas, is potentially curable if treated vigorously in an early phase.

2. The ability to measure the extent of the disease in skin and extracutaneous locations is becoming more precise, and this enables the clinician to define more accurately the type of treatment indicated.
3. Newer treatment programs are becoming available, which may make it possible to achieve a cure in early cases and significant prolongation of life in advanced cases.

Recently used topical agents in established MF include fluorinated corticosteroids (77, 78), nitrogen mustard (57, 76), nitrosourea compounds (79, 80), and topical cutaneous sensitizers (81, 82). Clinicians have long used topical corticosteroids in the treatment of premycotic eruptions, but studies have shown that fluorinated corticosteroid creams, when applied with occlusive techniques, will cause regression of mild to moderately advanced lesions (77, 78). This response is not based solely on the suppression of the inflammatory response, for serial biopsies during treatment have revealed the disappearance of both neoplastic and inflammatory cells (78). The adequacy of response is proportional to the concentration of the steroid, but the regression of lesions is temporary and there is ultimate recurrence.

Nitrogen mustard (HN2) was first employed for the topical therapy of MF more than 20 years ago and now assumes a major role in the treatment of MF confined to the skin (57, 76). Various treatment schedules have been employed; 10–40 mg of HN2 is dissolved in 50–60 ml of water, and the solution is applied to the entire body surface at intervals ranging from twice daily to two or three times a week. The time until total resolution of lesions varies from approximately one month to more than a year; the more extensive and more infiltrated lesions respond more slowly. The more extensive the disease, the less likely that a total remission can be achieved, although patients with the tumor stage of the disease have had complete regression of lesions. Heavily infiltrated plaques and tumors responding less adequately to topical HN2 have been directly injected with the drug (59). After regression of lesions, the frequency of HN2 applications can be reduced to once weekly for maintenance, and this is continued indefinitely. This form of therapy has produced an encouraging number of sustained remissions (57).

Topical HN2 has not been associated with systemic toxicity or bone-marrow depression, although a variable percentage of patients experience delayed hypersensitivity reactions to HN2 (83). Although patients with such hypersensitivity reactions often experience more rapid clearing of skin lesions, such reactions are not necessary for a satisfactory response (57). Hypersensitivity to HN2 has been managed by desensitizing the patient with the intravenous injection of minute amounts of HN2. Specific immune tolerance to HN2 also has been achieved before therapy is initiated (57).

Delayed hypersensitivity reactions in themselves have been found beneficial in stimulating regression of MF lesions. Patients with MF retain normal delayed hypersensitivity responses until late stages of the disease, and Ratner and associates (81) noted temporary regression of MF plaques after the application of 2, 4-dinitrochlorobenzene (DNCB) and S, 2-aminoethylisothiuronium (AET) to lesions in patients previously sensitized to these compounds by topical application. The purified protein derivative (PPD) of tuberculin has produced similar results (82), and the field of immunotherapy is currently under active investigation.

The nitrosoureas, another group of compounds used in systemic cancer chemotherapy, have produced encouraging results with topical application to MF plaques or to the entire body surface (79, 80). The more active compounds include 1,3-bis-(2-chloroethyl)-1-nitrosourea (BCNU) and 1-(2-chloroethyl)-3-cyclohexyl-1-nitrosourea (CCNU), and these have produced partial or complete regression of early to moderately infiltrated MF lesions. Bone marrow toxicity may occur, as may cutaneous telangiectasia and hyperpigmentation.

Various forms of radiotherapy have been used for many years in the treatment of MF, although local and whole-body orthovoltage irradiation has gradually given way to the use of electron beam therapy (3, 74, 84–86). Electrons generated at 2.5–3.5 MeV represent an effective mode of therapy of MF and have the advantages of (*a*) ease of whole-body irradiation at uniform dosage, (*b*) controlled limited penetration of the beam, (*c*) small integral dose of radiation delivered to the patient, and (*d*) limited bone marrow suppression. As in the case of topical therapy, use of the electron beam as the only means of therapy must be restricted to patients who do not have extracutaneous spread of MF.

In a currently employed program, delivery of 2000 to 3500 rads over a 40-day period has produced sustained total remission after a single course of treatment, which suggests the possibility of cure in patients with disease limited to the skin (3, 75). Success of therapy has been directly proportional to the total dose employed and inversely proportional to the extent of involvement.

Side effects of electron beam therapy have included cutaneous telangiectasia and poikiloderma at higher dose levels, and hyperpigmentation, xerosis, onycholysis, and epilation as less severe reactions. Bone marrow depression may also occur, but usually is transient with currently employed techniques.

Systemic chemotherapy has been reserved for patients with extracutaneous MF, and many have advanced disease at the time such therapy is initiated. At present, no mode of drug therapy has substantially lengthened the course of the disease (1), and remissions are usually short-lived and seldom exceed 6 months without maintenance therapy (16). With maintenance therapy, occasional patients may survive a year or longer (16).

Most therapeutic programs have utilized single drugs. Nitrogen mustard (87, 88) and cyclophosphamide (Cytoxan®) (89–91) have been used in relatively large numbers of patients, and good initial response rates have been achieved. Methotrexate (92) also has been widely used, with variable responses. Chlorambucil (Leukeran®) as a single agent has been disappointing in the treatment of MF.

Azaribine (93, 94) and bleomycin (Blenoxane®) (95–98) have been employed more recently and have produced some high response rates of varying duration. Use of systemic corticosteroids has been abandoned because of the brief duration of effectiveness.

Combined chemotherapy as employed in the treatment of other malignant lymphomas has not had widespread study in MF (1, 3, 16). The combination of nitrogen mustard, vincristine, procarbazine, and prednisone (MOPP) produced a favorable response in three of four patients (3), and a single patient remained clear of lesions for more than a year with the combination of cyclophosphamide, vincristine, and prednisone (COP) (99). Another patient, who had become resistant to a COP program and to other modalities, responded to bleomycin and remained clear of lesions with maintenance therapy with a combination of bleomycin, cyclophosphamide, and prednisone during an 8-month period of observation (97).

SÉZARY SYNDROME

The Sézary syndrome (erythroderma with circulating atypical mononuclear cells) has passed from engendering disbelief to being accepted as a clinical fact. There are four known typical and variant syndromes: pre-Sézary syndrome, Sézary syndrome, Sézary phenomenon, and T-cell leukemia. Reviews such as the Sézary symposium at the Mayo Clinic (100), the Sézary Grand Rounds at the NCI (7), and the Prosser White Oration (101) have summarized the clinical, cytologic, pathologic, and immunologic facts about this disease. It is now accepted that the Sézary cell is an atypical lymphocyte with a convoluted nucleus and an ample cytoplasm, often with vacuoles or pseudopods. In phase microscopy, cytochemistry, electron microscopy, and freeze-fracture studies, the cerebriform, folded nucleus is striking. Sézary syndrome is found in patients whose disease may begin as a chronic, spreading dermatitis or as an infiltrative erythema, eventually generalizing to a total erythroderma. Winkelmann (102) summarized the clinical findings in 28 cases, indicating that erythroderma, edema, and pruritus were present in all. Additional findings of alopecia, onychodystrophy, keratoderma, and pigmentary changes were often observed. Lymphadenopathy and hepatomegaly were benign when they occurred. Staging procedures in all cases, including bone marrow examination, liver and spleen scans, and laparotomy in five cases, were negative. Results of laboratory tests are noncontributory except for an occasional lymphocytosis. The diagnosis could be suggested by the benign dermal lymphocytic reticulosis observed as a bandlike infiltrate on skin biopsy in most cases (103). Pruniéras et al (104) have demonstrated with tritiated thymidine studies that these cells show DNA synthesis. Touch preparations produced clusters of atypical lymphocytes in half the cases studied.

In addition, the Sézary phenomenon, consisting of Sézary cytemia, may occur in mycosis fungoides on a transient basis or in lymphoma such as Hodgkin's disease as a reactive phase of the disease process in patients who may or may not have erythroderma. T-cell leukemia is most commonly observed in children, but is also seen in adults as a progression from the erythroderma of Sézary syndrome. This is exemplified by the NCI cases of erythroderma, lymphadenopathy, splenomegaly, bone marrow infiltration, elevated sedimentation rate, and markedly elevated white count of more than 100,000 cells/mm^3. Pre-Sézary syndrome may be recognized by low Sézary cell counts in the peripheral blood, but daily counts may reveal high pulses of Sézary cells (105).

Sézary syndrome may involute spontaneously, as it did in one of the Mayo Clinic patients (106). Lymphoma may occur, as it has in most of the NCI cases, but more rarely in cases reported in the literature and in only 4 of 28 cases at the Mayo Clinic.

The Mayo Clinic cases of lymphoma involved lymphosarcoma, reticulum-cell sarcoma, Hodgkin's disease, and myelomonocytic leukemia. Four additional patients were treated with chemotherapeutic drugs, and they died as a direct consequence of this therapy. Another major group died of cardiovascular and infectious complications of their illness, treatment, and old age without lymphoma. A large group of nine patients has been treated to remission with rest, nursing care, oral prednisone, and low-dose chlorambucil (2–6 mg/day). In the patients treated into remission, many positive patch tests were observed, indicating a high degree of active cutaneous delayed hypersensitivity. Three cases of photocontact dermatitis to salicylanilide are particularly striking. In three patients with atopic erythroderma, Sézary syndrome gradually developed while they were under observation, passing from pre-Sézary to Sézary syndrome. Normal lymphocytes in tissue culture may be stimulated by mitogen to resemble Sézary cells. Sézary syndrome can therefore develop from drug reactions, contact dermatitis, atopic dermatitis, or as a manifestation of these or other causes in patients with lymphoma.

The interest in the immunologic background of Sézary syndrome was heightened when a series of laboratories studied the cells and found an absence of receptors for immunoglobulins and complement and noted that the cells form rosettes with sheep RBCs (107). Massive numbers of T cells in the peripheral blood and skin were noted, and Sézary syndrome was termed a T-cell erythroderma. The cells were found to give a variable response to mitogens such as phytohemagglutinin and pokeweed, specific antigens, and mixed lymphocyte culture. The depression of response was most noteworthy in the NCI cases (108) and those of Braylin et al (109), perhaps related more to the lymphoma potential or reality in those cases. The defect of response in the NCI cases was not due to a circulating factor, but was a cellular property (7). It is recognized that there are patients with several clonal types of lymphoproliferative Sézary cells: the usual T cell, which is reactive and benign; the more malignant leukemic, unresponsive T lymphocyte; cells with properties of both T and B cells; and cells with null cell properties (110–112).

Carapeto et al (65) noted the increase in peripheral blood T-cell population, but also found a consistent decrease in B cells. A patient with atopic erythroderma whose disorder was observed to progress from pre-Sézary syndrome with 2–3% circulating Sézary cells to 30% Sézary cells showed a concomitant decrease in IgE levels, which indicated less B-cell activity (102). In leukopheresis, billions of T cells are removed, but the number of B cells increases. The balance of B and T cells may be as important as their individual values.

Crossen et al (107) noted cytogenetic changes in their case of Sézary syndrome, with hyperploidy and marker chromosomes. This finding was confirmed by Lutzner et al (113) and pointed to the selective proliferation of special clones of lymphocytes. Dewald et al (114) noted hypoploidy and marker chromosomes in a patient who had had Sézary syndrome for 9 years treated with considerable chemotherapy, and who died of infection without lymphoma. It is tempting to relate the chromosomal changes to malignancy, but cases such as the one cited above suggest caution. Many cases of Sézary syndrome at the Mayo Clinic and also now in Paris have been described in which there were normal chromosomes. In many cases with chromosome changes, malignancy has not developed.

Pathogenesis may be related to age-induced endogenous antigens or exogenous physical or chemically induced antigens, acording to Pruniéras (101). Chronic contact dermatitis may be the most common cause of the erythroderma, and this is supported by the multiple positive patch tests found in cases in remission. T-cell proliferation is also associated with circulating macrophage migration inhibition factor (115), and eight of nine reported cases of Sézary syndrome have shown migration inhibition factor activity (64). Tan et al (10) considered mycosis fungoides to be a disease of antigen persistence and therefore T-cell activity.

Many authors believe that Sézary syndrome is erythrodermic mycosis fungoides, and the presence of hyperconvoluted T cells in both states supports this hypothesis. It is admitted that the cell is present in banal dermatoses and that its presence is nonspecific. The evolution of three stages of mycosis fungoides contrasts with the dermatitis-erythroderma evolution of Sézary syndrome. Mycosis fungoides is epidermotropic, and Sézary syndrome is a dermotropic lymphoproliferative disease. It is perhaps wise to consider these two states as morphologically unique, T-cell-transitional, cutaneous responses to unknown antigens wherein the commitment to a specific course has not occurred. If we consider these as transitional lymphoreticular proliferations of the skin, then the spontaneous resolution, chronic persistent course, or progression to lymphoma may all be expected. Table 1 provides a partial list of benign, transitional, and malignant lymphoreticular diseases of the skin. Sézary syndrome is an immunoreactive state with a malignant potential that may be controlled by appropriate therapy. It offers a magnificent opportunity to study the T-cell response, the relation of T-cell and B-cell interaction in humans, prelymphomatous disease, and the transition to lymphoma.

Literature Cited

1. Epstein, E. H. Jr., Levin, D. L., Croft, J. D. Jr., Lutzner, M. A. 1972. Mycosis fungoides: survival, prognostic features, response to therapy, and autopsy findings. *Medicine (Baltimore)* 51: 61–72
2. Fuks, Z. Y., Bagshaw, M. A., Farber, E. M. 1973. Prognostic signs and the management of the mycosis fungoides. *Cancer* 32:1385–95
3. Levi, J. A., Wiernik, P. H. 1975. Management of mycosis fungoides: current status and future prospects. *Medicine (Baltimore)* 54:73–88
4. Symmers, D. 1932. Mycosis fungoides as a clinical and pathologic nonexistent. *Arch. Dermatol. Syphilol.* 25:1–5
5. Reed, R. J., Cummings, C. E. 1966. Malignant reticulosis and related conditions of the skin: a reconsideration of mycosis fungoides. *Cancer* 19:1231–47
6. Clendenning, W. E. 1971. Mycosis fungoides. In *Modern Trends in Dermatology—4*, ed. P. Borrie, pp. 214–37. New York: Appleton-Century-Crofts. 292 pp.
7. Lutzner, M. A., Edelson, R., Schein, P., Green, I., Kirkpatrick, C., Ahmed, A. 1975. Cutaneous T-cell lymphomas: the Sézary syndrome, mycosis fungoides, and related disorders. *Ann. Intern. Med.* 83:534–52
8. Samman, P. D. 1972. The natural history of parapsoriasis en plaques (chronic superficial dermatitis) and prereticulotic poikiloderma. *Br. J. Dermatol.* 87:405–11
9. Samman, P. D. 1964. Survey of reticuloses and premycotic eruptions: a preliminary report. *Br. J. Dermatol.* 76:1–9
10. Tan, R. S.-H., Butterworth, C. M., McLaughlin, H., Malka, S., Samman, P. D. 1974. Mycosis fungoides: a disease of antigen persistence. *Br. J. Dermatol.* 91:607–16
11. Winkelmann, R. K. 1972. Benign and malignant lymphocytic reticuloses. *Excerpta Med. Int. Congr. Ser.* No. 248, pp. 24–25 (Abstr.)

12. Schwartz, R. S., Beldotti, L. 1965. Malignant lymphomas following allogenic disease: transition from an immunological to a neoplastic disorder. *Science* 149:1511–14
13. Bluefarb, S. M. 1955. Is mycosis fungoides an entity? *Arch. Dermatol. Syphilol.* 71:293–302
14. Degos, R., Civatte, J., Touraine, R., Ossipovski, B., Gonzalez-Mendoza, A. 1965. Confrontation anatomo-clinique de 129 hémoréticulopathies malignes cutanées. *Ann. Dermatol. Syphiligr.* 92:121–27
15. Bluefarb, S. M. 1960. *Cutaneous Manifestations of the Reticuloendothelial Granulomas.* Springfield, Ill: Thomas. 442 pp.
16. Rappaport, H., Thomas, L. B. 1974. Mycosis fungoides: the pathology of extracutaneous involvement. *Cancer* 34: 1198–1229
17. Long, J. C., Mihm, M. C. 1974. Mycosis fungoides with extracutaneous dissemination: a distinct clinicopathologic entity. *Cancer* 34:1745–55
18. Allen, A. C. 1967. *The Skin: A Clinicopathological Treatise*, Pts. 1, 2. New York: Grune & Stratton. 591 pp., 591 pp. 2nd ed.
19. Gall, E. A. 1955. Enigmas in lymphoma: reticulum cell sarcoma and mycosis fungoides. *Minn. Med.* 38:674–81; 705
20. Montgomery, H. 1967. *Dermatopathology*, Vols. 1, 2. New York: Harper & Row. 678 pp., 624 pp.
21. Montgomery, H., Winkelmann, R. K. 1968. Histopathologic and hematologic changes in malignant lymphomas and reticuloses. In *International Congress of Dermatology, 13th,* pp. 707–12. New York: Springer. 1598 pp.
22. Orfanos, C., Stüttgen, G. 1963. Elektronenmikroskopische Beobachtungen bei der Mycosis fungoides: Nachweis von cytoplasmatischen "Zelleinschlüssen." *Arch. Klin. Exp. Dermatol.* 215: 438–60
23. Bonneau, H., Cesarini, J. -P. 1967. Étude ultrastructurale d'un cas de mycosis fongoïde. *Bull. Soc. Fr. Dermatol. Syphiligr.* 74:612–17
24. Lutzner, M. A., Jordan, H. W. 1968. The ultrastructure of an abnormal cell in Sézary's syndrome. *Blood* 31:719–26
25. Lutzner, M. A., Hobbs, J. W., Horvath, P. 1971. Ultrastructure of abnormal cells: in Sézary syndrome, mycosis fungoides, and parapsoriasis en plaque. *Arch. Dermatol.* 103:375–86
26. Fisher, E. R., Horvat, B. L., Wechsler, H. L. 1972. Ultrastructural features of mycosis fungoides. *Am. J. Clin. Pathol.* 58:99–110
27. Sandbank, M. 1971. Mycosis fungoides: ultrastructural study with demonstration of atypical cells and nuclear bodies. *Arch. Dermatol.* 103:206–14
28. Flaxman, B. A., Zelazny, G., Van Scott, E. J. 1971. Nonspecificity of characteristic cells in mycosis fungoides. *Arch. Dermatol.* 104:141–47
29. Rosas-Uribe, A., Variakojis, D., Molnar, Z., Rappaport, H. 1974. Mycosis fungoides: an ultrastructural study. *Cancer* 34:634–45
30. Burg, G., Braun-Falco, O. 1974. Fortschritte in der cytochemischen Differenzierung von Infiltratzellen in der haut. *Hautarzt* 25:1–8
31. Burg, G., Braun-Falco, O. 1974. Qualitative und quantitative Aspekte der cellulären Reaktion in Haut und Blut bei Mycosis fungoides: eine cytochemische Studie. *Hautarzt* 25:178–87
32. Vesper, L. J., Winkelmann, R. K., Hargraves, M. M. 1971. The mycosis fungoides cell: the skin window in mycosis fungoides. *Br. J. Dermatol.* 84:54–65
33. Meuret, G., Lowka, K., Brand, E. T., Kalkoff, K. W. 1973. Erythrodermia, mycosis fungoides, skin reticulosis: autonomous disorders of the monocytopoietic macrophage system. *Acta Haematol.* 50:329–39
34. Seitz, L. E., Golitz, L. E., Weston, W. L. 1976. Defective chemotaxis on mononuclear leukocytes in mycosis fungoides. *Clin. Res.* 24:98A (Abstr.)
35. Goldschmidt, H., Thew, M. A. 1972. Exfoliative cytology of psoriasis and other common dermatoses: quantitative analysis of parakeratotic horny cells in 266 patients. *Arch. Dermatol.* 106: 476–83
36. Robinowitz, B., Roenigk, H. H. Jr., Smith, M. M. 1975. Sea-blue histiocytes in mycosis fungoides. *Arch. Dermatol.* 111:1165–67
37. Brehmer-Andersson, E., Brunk, U. 1967. Tape-stripping method for cytological diagnosis of mycosis fungoides. *Acta Derm. Venereol.* 47:177–80
38. Grosshans, E., Hee, P., Basset, A., Maleville, J. 1973. Le maladie de Woringer et Kolopp (réticulose cutanée épidermotrope, "réticulose pagétoide"). *Arch. Belg. Dermatol.* 29:195–203
39. Braun-Falco, O., Marghescu, S., Wolff, H. H. 1973. Pagetoide Reticulose: Morbus Woringer-Kolopp. *Hautarzt* 24:11–21

40. Revuz, J., Boulvin, F., Weshsler, J., Touraine, R. 1975. Une pathologie pour la cellule de Merkel? La "reticulose" de Woringer et Kolopp. *Nouv. Presse Med.* 4:1878

41. Van Vloten, W. A., van Duijn, P., Schaberg, A. 1974. Cytodiagnostic use of Feulgen-DNA measurements in cell imprints from the skin of patients with mycosis fungoides. *Br. J. Dermatol.* 91:365–71

42. Erkman-Balis, B., Rappaport, H. 1974. Cytogenetic studies in mycosis fungoides. *Cancer* 34:626–33

43. Lever, W. F., Schaumburg-Lever, G. 1975. *Histopathology of the Skin,* pp. 696–703. Philadelphia: J. B. Lippincott. 793 pp. 5th ed.

44. Kim, R., Winkelmann, R. K., Dockerty, M. 1963. Reticulum cell sarcoma of the skin. *Cancer* 16:646–55

45. Petrozzi, J. W., Raque, C. J., Goldschmidt, H. 1971. Malignant lymphoma, reticulum cell type: ultrastructural and cytological demonstration of Lutzner cells. *Arch. Dermatol.* 104: 38–44

46. Ackerman, A. B., Breza, T. S., Capland, L. 1974. Spongiotic simulants of mycosis fungoides. *Arch. Dermatol.* 109:218–20

47. Macaulay, W. L. 1968. Lymphomatoid papulosis: a continuing self-healing eruption, clinically benign–histologically malignant. *Arch. Dermatol.* 97: 23–30

48. Black, M. M., Wilson Jones, E. 1972. "Lymphomatoid" pityriasis lichenoides: a variant with histological features simulating a lymphoma; a clinical and histopathological study of 15 cases with details of long term follow-up. *Br. J. Dermatol.* 86:329–47

49. Valentino, L. A., Helwig, E. B. 1973. Lymphomatoid papulosis. *Arch. Pathol.* 96:409–16

50. Block, J. B., Edgcomb, J., Eisen, A., Van Scott, E. J. 1963. Mycosis fungoides: natural history and aspects of its relationship to other malignant lymphomas. *Am. J. Med.* 34: 228–35

51. Cyr, D. P., Geokas, M. C., Worsley, G. H. 1966. Mycosis fungoides: hematologic findings and terminal course. *Arch. Dermatol.* 94:558–73

52. Cawley, E. P., Curtis, A. C., Leach, J. E. K. 1951. Is mycosis fungoides a reticuloendothelial neoplastic entity? *Arch. Dermatol. Syphilol.* 64:255–71

53. Variakojis, D., Rosas-Uribe, A., Rappaport, H. 1974. Mycosis fungoides: pathologic findings in staging laparotomies. *Cancer* 33:1589–1600

54. Fuks, Z. Y., Castellino, R. A., Carmel, J. A., Farber, E. M., Bagshaw, M. A. 1974. Lymphography in mycosis fungoides. *Cancer* 34:106–12

55. Blaylock, W. K., Clendenning, W. E., Carbone, P. P., Van Scott, E. J. 1966. Normal immunologic reactivity in patients with the lymphoma mycosis fungoides. *Cancer* 19:233–36

56. Clendenning, W. E., Van Scott, E. J. 1965. Skin autografts and homografts in patients with the lymphoma mycosis fungoides. *Cancer Res.* 25:1844–53

57. Van Scott, E. J., Kalmanson, J. D. 1973. Complete remissions of mycosis fungoides lymphoma induced by topical nitrogen mustard (HN2): control of delayed hypersensitivity to HN2 by desensitization and by induction of specific immunologic tolerance. *Cancer* 32:18–30

58. Lai A Fat, R. F. M., Cormane, R. H., Van Furth, R. 1974. An immunohistopathological study on the synthesis of immunoglobulins and complement in normal and pathological skin and the adjacent mucous membranes. *Br. J. Dermatol.* 90:123–36

59. Tuffanelli, D. L. 1975. Cutaneous immunopathology: recent observations. *J. Invest. Dermatol.* 65:143–53

60. Winkelmann, R. K. 1973. T cell erythroderma (Sezary syndrome). *Arch. Dermatol.* 108:205–6

61. Robinowitz, B. N., Noguclie, S., Roenigk, H. H. 1975. Tumor cell characteristics in mycosis fungoides. Presented at Meet. Eur. Soc. Dermatol. Res.-Soc. Invest. Dermatol., Amsterdam, June 9–13

62. Clendenning, W. E., Brecher, G., Van Scott, E. J. 1964. Mycosis fungoides: relationship to malignant cutaneous reticulosis and the Sézary syndrome. *Arch. Dermatol.* 89:785–92

63. Langner, A., Gliński, W., Pawińska, M., Obałek, S. 1975. Lymphocyte transformation in mycosis fungoides. *Arch. Dermatol. Forsch.* 251:249–57

64. Umbert, P., Belcher, R. W., Winkelmann, R. K. 1976. Macrophage inhibitor factor (MIF) in cutaneous lymphoproliferative diseases. *Br. J. Dermatol.* 95:475–80

65. Carapeto, F. J., Stobo, J. D., Jordon, R. E., Winkelmann, R. K. 1975. Población linfocitaria (T y B linfocitos) en sangre periférica de controles sanos y algunas

dermatósis (resultados previos). *Actas Dermo-Sifiliogr.* 66:243–64

66. Kawacki, A., Ogawa, H., Mori, S., Hayaski, T. 1971. Serum immunoglobulins in mycosis fungoides. *Jpn. J. Dermatol. Ser. B* 81:498–506
67. Herrmann, W. P., Gartmann, H. 1974. Immunelektrophoretische Befunde bei malignen Retikulosen und Retikulogranulomatosen. *Z. Hautkr.* 49:661–66
68. Jones, S. E. 1975. Non-Hodgkin lymphomas. *J. Am. Med. Assoc.* 234:633–38
69. Rappaport, H., Winter, W. J., Hicks, E. B. 1956. Follicular lymphoma: a reevaluation of its position in the scheme of malignant lymphoma, based on a survey of 253 cases. *Cancer* 9:792–821
70. Carbone, P. P., Kaplan, H. S., Musshoff, K., Smithers, D. W., Tubiana, M. 1971. Report of the committee on Hodgkin's disease staging classification. *Cancer Res.* 31:1860–61
71. Winkelmann, R. K. 1974. Lymphocytic reticulosis and Sezary syndrome. *Jpn. J. Dermatol.* 1:1–13
72. Griem, M. L., Moran, E. M., Ferguson, D. J., Mettler, F. A., Griem, S. F. 1975. Staging procedures in mycosis fungoides. *Br. J. Cancer* 31: Suppl. 2, pp. 362–67
73. Escovitz, E. S., Soulen, R. L., Van Scott, E. J., Kalmanson, J. D., Barry, W. E. 1974. Mycosis fungoides: a lymphographic assessment. *Radiology* 112: 23–27
74. Haynes, H. A., Van Scott, E. J. 1968. Therapy of mycosis fungoides. *Prog. Dermatol., Dermatol. Found. Bull.* 3:1–4
75. Fuks, Z., Bagshaw, M. A. 1971. Total-skin electron treatment of mycosis fungoides. *Radiology* 100:145–50
76. Van Scott, E. J., Winters, P. L. 1970. Responses of mycosis fungoides to intensive external treatment with nitrogen mustard. *Arch. Dermatol.* 102:507–14
77. Farber, E. M., Cox, A. J., Steinberg, J., McClintock, R. P. 1966. Therapy of mycosis fungoides with topically applied fluocinolone acetonide under occlusive dressing. *Cancer* 19:237–45
78. Farber, E. M., Zackheim, H. S., McClintock, R. P., Cox, A. J. Jr. 1968. Treatment of mycosis fungoides: with various strengths of fluocinolone acetonide cream. *Arch. Dermatol.* 97:165–72
79. Zackheim, H. S. 1972. Treatment of mycosis fungoides with topical nitrosourea compounds. *Arch. Dermatol.* 106:177–82
80. Zackheim, H. S., Epstein, E. H. Jr. 1975. Treatment of mycosis fungoides with topical nitrosourea compounds: further studies. *Arch. Dermatol.* 111: 1564–70
81. Ratner, A. C., Waldorf, D. S., Van Scott, E. J. 1968. Alterations of lesions of mycosis fungoides lymphoma by direct imposition of delayed hypersensitivity reactions. *Cancer* 21:83–88
82. Klein, E., Burgess, G. H., Helm, F. 1973. Neoplasms of the skin. In *Cancer Medicine,* ed. J. F. Holland, E. Frei III, pp. 1789–1822. Philadelphia: Lea & Febiger. 2018 pp.
83. Waldorf, D. S., Haynes, H. A., Van Scott, E. J. 1967. Cutaneous hypersensitivity and desensitization to mechlorethamine in patients with mycosis fungoides lymphoma. *Ann. Intern. Med.* 67:282–90
84. Grollman, J. H. Jr., Bierman, S. M., Ottoman, R. E., Morgan, J. E., Horns, J. 1966. Total-skin electron-beam therapy of lymphoma cutis and generalized psoriasis: clinical experiences and adverse reactions. *Radiology* 87:908–15
85. D'Angio, G. J., Nisce, L. Z., Ho Kim, J. 1975. Weekly total skin electron beam therapy for mycosis fungoides and other cutaneous lymphomata: further experience. *Br. J. Cancer* 31: Suppl. 2, pp. 379–85
86. Szur, L. 1975. The treatment of mycosis fungoides and related conditions with particular emphasis on electron therapy. *Br. J. Cancer* 31: Suppl. 2, pp. 368–78
87. Henstell, H. H., Tober, J. N. 1947. Treatment of mycosis fungoides with nitrogen mustard. *J. Invest. Dermatol.* 8:183–88
88. Kierland, R. R., Watkins, C. H., Shullenberger, C. C. 1947. The use of nitrogen mustard in the treatment of mycosis fungoides. *J. Invest. Dermatol.* 9:195–201
89. Abele, D. C., Dobson, R. L. 1960. The treatment of mycosis fungoides with a new agent, cyclophosphamide (Cytoxan). *Arch. Dermatol.* 82:725–31
90. Van Scott, E. J., Auerbach, R., Clendenning, W. E. 1962. Treatment of mycosis fungoides with cyclophosphamide. *Arch. Dermatol.* 85:499–501
91. Maguire, A. 1968. Treatment of mycosis fungoides with cyclophosphamide and chlorpromazine. *Br. J. Dermatol.* 80:54–57
92. Wright, J. C., Lyons, M. M., Walker, D. G., Golomb, F. M., Gumport, S. L.,

Medrek, T. J. 1964. Observations on the use of cancer chemotherapeutic agents in patients with mycosis fungoides. *Cancer* 17:1045–62

93. Calabresi, P., Turner, R. W. 1966. Beneficial effects of triacetyl azauridine in psoriasis and mycosis fungoides. *Ann. Intern. Med.* 64:352–71
94. McDonald, C. J., Calabresi, P. 1971. Azaribine for mycosis fungoides. *Arch. Dermatol.* 103:158–67
95. de Bast, C., Moriame, N., Wanet, J., Ledoux, M., Achten, G., Kenis, Y. 1971. Bleomycin in mycosis fungoides and reticulum cell lymphoma. *Arch. Dermatol.* 104:508–12
96. European Organization for Research on the Treatment of Cancer, Co-operative Group for Leukaemia and Reticuloses. 1972. Bleomycin in the recticuloses. *Br. Med. J.* 1:285–86
97. Spigel, S. C., Coltman, C. A. Jr. 1973. Therapy of mycosis fungoides with bleomycin. *Cancer* 32:767–70
98. Van Vloten, W. A., Polano, M. K. 1975. Bleomycin therapy in mycosis fungoides. *Dermatologica* 150:50–57
99. Leavell, U. W. Jr., Mersack, I. P., Johnson, J. L., Butler, D. 1975. Mycosis fungoides: case report *J. Ky. Med. Assoc.* 73:139–42
100. Winkelmann, R. K. 1974. The Sezary symposium. *Mayo Clin. Proc.* 49: 515–92
101. Pruniéras, M. 1975. The Sezary's syndrome. *Trans. St. John's Hosp. Dermatol. Soc.* 61:1–10
102. Winkelmann, R. K. 1974. Clinical studies of T-cell erythroderma in the Sezary syndrome. *Mayo Clin. Proc.* 49:519–25
103. Holdaway, D., Winkelmann, R. K. 1974. Histopathology of Sezary syndrome. *Mayo Clin. Proc.* 49:541–47
104. Pruniéras, M., Saglier, I., Grupper, C., Durepaire, R., Flandrin, G. 1975. Le syndrome de Sézary. *Pathol. Biol.* 23:553–65
105. Heaphy, M. R., Winkelmann, R. K. 1976. Variations in peripheral Sézary cell count. *Arch. Dermatol.* 112:560
106. Winkelmann, R. K., Perry, H. O., Muller, S. A., Schroeter, A. L., Jordon, R. E., Rogers, R. S. III. 1974. Treatment of Sezary syndrome. *Mayo Clin. Proc.* 49:590–92
107. Crossen, P. E., Mellor, J. E. L., Finley, A. G., Ravich, R. B. M., Vincent, P. C., Gunz, F. W. 1971. The Sézary syndrome: cytogenetic studies and identification of the Sézary cell as an abnormal lymphocyte. *Am. J. Med.* 50:24–34
108. Edelson, R. L., Lutzner, M. A., Kirkpatrick, C. H., Shevach, E. M., Green, I. 1974. Morphologic and functional properties of the atypical T lymphocytes of the Sezary syndrome. *Mayo Clin. Proc.* 49:558–66
109. Braylin, R., Variakojis, D., Yachnin, S. 1976. The Sezary syndrome lymphoid cell: abnormal surface properties and nitrogen responsiveness. *J. Clin. Invest.* In press
110. Zucker-Franklin, D., Melton, J. W. III, Quagliata, F. 1973. Variants of chronic lymphocytic leukemia with cells having surface properties of T-lymphocytes. *J. Clin. Invest.* 52:92a (Abstr.)
111. Flandrin, G., Brouet, J.-C. 1974. The Sezary cell: cytologic, cytochemical, and immunologic studies. *Mayo Clin. Proc.* 49:575–83
112. Shevach, E., Edelson, R., Frank, M., Lutzner, M., Green, I. 1974. A human leukemia cell with both B and T cell surface receptors. *Proc. Natl. Acad. Sci. USA* 71:863–66
113. Lutzner, M. A., Emerit, I., Durepaire, R., Flandrin, G., Grupper, C., Pruniéras, M. 1973. Cytogenetic, cytophotometric, and ultrastructural study of large cerebriform cells of the Sézary syndrome and description of a small-cell variant. *J. Natl. Cancer Inst.* 50:1145–62
114. Dewald, G., Spurbeck, J. L., Vitek, H. A. 1974. Chromosomes in a patient with the Sezary syndrome. *Mayo Clin. Proc.* 49:553–57
115. Yoshida, T., Edelson, R., Cohen, S., Green, I. 1975. Migration inhibitory activity in serum and cell supernatants in patients with Sezary syndrome. *J. Immunol.* 114:915–18

Ann. Rev. Med. 1977. 28:271–89
Copyright © 1977 by Annual Reviews Inc. All rights reserved

CRITIQUE OF CORONARY ARTERY BYPASS SURGERY

◆7242

Charles B. Mullins, M.D. and Kirk Lipscomb, M.D.
Department of Internal Medicine, Division of Cardiology, University of Texas
Health Science Center at Dallas and the Dallas Veterans Administration Hospital,
Dallas, Texas 75235

INTRODUCTION

Coronary artery disease afflicts 5% of the United States adult population; 3.1 million Americans between ages 18 and 79 definitely have coronary artery disease and another 2.4 million are suspected to have the disease. There are in excess of one million myocardial infarctions per year in our country, with a mortality rate approaching 40%.

There is little wonder that coronary artery bypass surgery, since its inception 10 years ago, has rapidly risen to its present position as one of the most frequent surgical procedures in this country. Indications are constantly changing, due to continued analysis of the results of surgery in terms of patient morbidity, mortality, and relief of symptoms.

Perhaps more importantly, this surgery has stimulated an intensive study into the natural history of patients with coronary artery disease expressed as angina pectoris or myocardial infarction, and the knowledge gained from this natural history of coronary disease has itself dictated changes in the indications for coronary bypass surgery.

Surprisingly, few objective, well-designed, randomized studies have been reported; thus current information must be obtained by the accumulation of a large number of often incomplete or not ideally designed reports.

This review attempts to assimilate the current data available on coronary bypass surgery and to analyze it as objectively as possible.

OPERATIVE PROCEDURE

Bypass of the diseased proximal coronary artery is now most often accomplished by grafting a saphenous vein from the proximal aorta to the distal coronary artery, or less commonly by grafting the distal end of the internal mammary artery to the distal coronary artery. The internal mammary is usually reserved for low-flow situations where vein-graft patency might be threatened. The advantages of the internal mammary are its high patency rate and the likelihood that long-term prognosis is better with an arterial conduit.

Operative Mortality

Overall surgical mortality for coronary bypass is now less than 5% and, as expected, the mortality varies between institutions (1–20). In those institutions reporting their experience over a period of years, the mortality decreases with time. (2, 7–9, 14, 19). This drop in mortality probably reflects increased technical expertise, but more importantly, greater experience in patient selection. Undoubtedly the most important factor in predicting the risk of surgical mortality is resting left-ventricular function. Second in importance is the presence of left main coronary disease. Unstable angina, number of vessels diseased, previous propranolol therapy, age, and diabetes all contribute much less to operative mortality or are insignificant.

The significant effect of poor resting ventricular function on operative mortality is illustrated in Table 1 (12, 21–23). The parameter of measurement of ventricular function in these studies is the ejection fraction (the ratio of stroke volume to end-distolic volume). The lower limit of normal for the ejection fraction is generally accepted as 50%.

Left main coronary disease is found in about 4% of patients with coronary disease (24–28), and when present contributes significantly to an increase in operative mortality. Early mortality figures were as high as 31% (23, 25); but with improved surgical technology (29), this figure has decreased to about 10% (24, 26–28, 30–34).

In earlier reports, surgery in patients with unstable angina resulted in a higher mortality than surgery for stable angina, (35, 36). However, since the initiation of

Table 1 Surgical mortality related to ventricular ejection fraction

Ejection fraction (%)	Operative mortality (%)
Seattle VA Hospital (12)	
< 33	33
> 33	3
Peter Bent Brigham Hospital (21)	
< 50	35
> 50	3
Penn State University Hospital (22)	
12-30	25
31-40	12
41-50	4
> 50	4
Duke University Hospital (23)	
< 25	55
> 25	4

the practice of stabilizing patients medically prior to surgery, this mortality has dropped (35, 36–39), so that now it is comparable to the mortality for stable angina (37–44).

Generally, the number of diseased coronary arteries or the number of grafts placed has little influence on mortality (3, 5, 8, 9, 12, 19, 20, 23, 45).

Older patients, especially those over 70 years of age, have a higher surgical mortality (14, 20, 46, 47); however, in at least one study this higher mortality was attributed to a much higher incidence of poor ventricular function in this older group (46). Diabetes does not appear to contribute significantly to surgical risk (48, 49), but may influence patency rates of the grafts.

Surgery in the immediate post–myocardial infarction period carries a mortality of 20–100% (50–52) and certainly affects initial surgical mortality. The relationship between time after infarct and surgical mortality in one study is shown in Table 2 (53). Since surgery in the post-infarct period is usually performed only on severely ill patients, this high mortality probably reflects the result of surgery on patients with infarct-related complications, rather than the temporal effect of the infarct itself.

There appears little indication for coronary-artery surgery in the setting of an uncomplicated acute myocardial infarction until the survival with coronary artery surgery increases to above the 85% survival in coronary-care units (CCUs) in patients with acute myocardial infarction.

Patients in cardiogenic shock due to acute myocardial infarction have been managed with intraaortic balloon assistance, followed by catheterization and coronary bypass surgery if they remain balloon-dependent. This was partly stimulated by the report of low operative mortality in patients treated surgically immediately after uncomplicated infarction (54) and an initial report of increased survival in patients treated surgically for shock (55). However, later reports overwhelmingly suggest that the operative mortality is inordinately high (51, 56, 57). Surgery should probably not be undertaken in the majority of these patients.

Operative Morbidity

Myocardial infarction occurs in about 10% of coronary-bypass operations with the range reported to be 4–21% (2, 5, 10, 13, 14, 16, 47, 58–67). Despite the popular thought that infarction during surgery is more benign than the usual nonsurgical infarct, the reported perioperative infarction mortality of 20–40% (5, 6, 45, 47, 65,

Table 2 Surgical mortality related to acute myocardial infarction

Days post infarct	Surgical mortality (%)
0–7	38
8–30	16
31–60	6
> 60	6
No infarct	4

66) suggests that it carries the same significance. Myocardial infarction is the cause of most immediate surgical deaths and is associated with almost all deaths in the 1–30-day postoperative period (13, 60, 68). Although infarction is definitely associated with a higher than usual incidence of graft occlusion, about as many patent grafts as occluded grafts are found to the infarcted area (16, 68), and about 25% of infarcts occur in areas supplied by ungrafted vessels (62, 68). Graft occlusion does not necessarily cause infarction—in one study only 15% of graft occlusions resulted in infarction (69).

Bypass grafting is associated with an increase in the progression of proximal coronary artery obstruction during the following six to twelve months in about 60% of cases (range:20–73%) compared to 15% (range:10–23%) in ungrafted arteries in the same patients. Proximal occlusion occurs in about 50% (range:20–63%) compared to 5% (range:2–8%) in ungrafted arteries (70–77). Although the rate of progression is slightly lower in arteries with nonpatent grafts (71, 72, 74), the combination of proximal vessel and graft occlusion does cause a net decrease in perfusion to some segments after surgery (74). When a graft occludes in the immediate postoperative period, proximal-artery occlusion is very common, occurring in 12 out of 13 cases in one study (74). This combination of obstructed graft and proximal-artery occlusion is frequent in early postoperative deaths (68). Grafting affects the atherosclerotic process little if at all in the artery distal to the anastomosis (70, 73, 74, 76).

In general, 10% of vein grafts occlude within two weeks of surgery and an additional 10–20% occlude between two weeks and one year for a one-year patency rate of 75% (range 60–86%) (2, 9, 14, 45, 57, 66, 67, 78, 79). After one year occlusion is rare and, when it does occur, can be predicted by severe stenosis at the one-year catheterization study (78, 80). Early occlusion is due to thrombosis (68) and late occlusion, to intimal fibrosis (68, 81–85). Atherosclerosis in the vein graft has been reported, but has not yet been reported to cause obstruction (68, 82, 85). The frequency of early occlusion is strongly dependent on the size of the coronary artery distal to the obstruction and the amount of myocardium it perfuses (86, 87). Graft flow measured at surgery is a predictor of early occlusion, with 40 ml/min commonly considered the separation value between adequate and compromised flow (79). In one study, all grafts with flows less than 25 ml/min were occluded two weeks later, while grafts with flows over 45 ml/min were patent (88). The cause of occlusion between one month and one year, i.e. the cause of intimal fibrosis, is unknown. The occlusion rate of 10–20% seems related little if at all to the factors responsible for early occlusion (88). Neither early nor late occlusion appears related to the specific coronary artery grafted, or to associated diabetes, hypertension, or occlusion in another graft in the same patient (49, 79). The patency of internal mammary-coronary artery anastomoses is higher and approaches 97% (17, 18).

SURVIVAL

The most important unanswered question regarding coronary-artery surgery is its effect on patients' long-term survival. Except in the case of left main disease, no adequately designed study has yet shown a statistically significant influence of surgery on long-term survival. However, inadequately designed studies have suggested that survival is prolonged after surgery when compared with other studies of medically treated patients. Controversy now exists over whether these data are adequate to recommend coronary-bypass surgery for the purpose of improving survival (89, 90). Because of the importance of this decision, the studies relative to surgical versus medical survival are reviewed in Table 3 (2, 9, 45, 91–96). It is readily apparent that the least-controlled studies show the most significant improvement. It is further apparent that the best-designed studies have the fewest patients, making it necessary to have a relatively large difference between surgical and medical mortality to be statistically significant. The results of the first five studies along with the medical series to which they are usually compared are shown in Table 4 (2, 9, 45, 91, 92). The mean follow-up was 2 years except at Methodist (Houston), where it was 1 year and Stanford, where it was 2.5 years. Operative mortality, which is included in the annual mortality, was 3% at Oregon and Cleveland; 7% at St. Luke's (Houston); and 5% at Stanford and Methodist (Houston). In this table and all subsequent tables, annual mortality assumes a linear relationship for each year and is based on the original cohort. These surgical results are most commonly compared to the natural-history data reported by Reeves and associates (97) and the Cleveland Clinic (98). These natural-history studies came chiefly from the time period before surgery. With the exception of the Cleveland data, the surgical and medical groups came from different hospitals and were categorized by different observers. In the Cleveland study the groups came from different time periods. For comparison, the Cleveland group placed equivalent numbers of patients in the surgical and medical groups according to number of diseased vessels. The Oregon and Stanford studies subdivided their patients for comparison. Although the St. Luke's and Methodist studies were not well categorized, most patients had multivessel disease.

Table 4 shows an apparent favorable influence of surgery on survival and is the basis for most of the statements to this effect. There are several objections to this conclusion. The data contained in Reeves' summary (97) is basically the natural

Table 3 Reported series of medical vs surgical treatment of coronary artery disease with analysis of the design of each series

Hospital	Number of surgical patients	Control (same hospital)	Control (same hospital same time)	Prospective randomized	Surgical improvement in survival
Oregon (9)	532				+
Stanford (91)	350				+
St. Luke's (Houston) (92)	1492				+
Methodist (Houston) (45)	480				+
Cleveland (2)	1000	+			+
Duke (93)	379	+	+		±
Long Beach VA (94, 95)	20	+	+	+	
Houston VA (96)	35	+	+	+	

Table 4 Five surgical series compared with the expected medical mortality analyzed with respect to coronary anatomy

	Annual mortality				
	Overall	1-vessel disease	2-vessel disease	2- & 3-vessel disease	3-vessel disease
Surgical Series					
Oregon (9)	3	1	3		5
Stanford (91)	5	2		6	
St. Luke's Hosp (92)	5				
Methodist Houston (45)	8				
Cleveland (2)	5				
Medical Series (97, 98)		2	7		11

survival of patients with coronary artery disease relative to the number of vessels diseased. This subdivision of expected mortality according to number of diseased vessels has certainly contributed to the comparability of medical and surgical survivorship. However, to use these data as the basis for a control group from another hospital assumes no variance between observers. The natural history of the medical control groups comes from an earlier time period when cineangiographic equipment had less resolution than in the later period when the surgical patients were catheterized, and thus may not be true anatomical comparisons. Also, these studies do not evaluate distal-vessel disease, a condition that classifies a patient as a poor surgical candidate and that would influence the inclusion of patients with this finding in any surgical series.

Another problem with comparing the uncontrolled data on surgical survival with the medical series is that ventricular function was not systematically evaluated. As shown in Table 5 (93, 99, 100), ventricular function is probably the single most important factor in predicting natural survival. Since poor ventricular function is well known to affect operative mortality adversely, this bias was present during the selection of candidates for surgery and probably results in some patients with poor ventricular function not being included in the surgical series.

There are three studies comparing surgical to medical mortality that are either well-controlled or randomized. These studies are summarized in Table 6 (93–96). At Duke the risk factors in a nonrandomized group of medical and surgical patients were compared and if unequal, statistically cancelled. As shown in Table 6, there was no significant difference in mortality. The statistician for this study estimated that if this comparison of mortality rates is valid, it would take approximately 100,000 patients to show significance. At the Long Beach VA Hospital, 40 patients were matched both as to number of diseased vessels and ventricular function, then randomly assigned to surgery or medicine. There was no significant difference in mortality. In the Houston VA Hospital prospectively randomized study, over half of the patients had three-vessel disease. Mean follow-up time was 28 months. Surgical mortality was lower, but the difference was insignificant.

Table 5 Surgical mortality related to anatomical coronary disease and ventricular function

	Annual mortality (%)			
	Ventricular function			
Hospital	Normal	Moderately abnormal	Very abnormal	Method of determination of ventricular function
Duke (93)				angio[a],$A\text{-}V_{O_2}\Delta$[b]
1-vessel	2	0	14	
2-vessel	0	9	20	
3-vessel	5	12	20	
Peter Bent Brigham (99)				angio[a] ejection fraction
1-vessel	1	19	43	
3-vessel	10	29	48	
Alabama (100)				chest X ray
1-vessel	0	3		
2-vessel	3	21		
3-vessel	14	24		

[a] angio = cineangiographic
[b] $A\text{-}V_{O_2}$ Δ = arterial-venous oxygen difference

Several subgroups deserve special attention. Left main coronary disease is more lethal than other types of coronary disease. Annual mortality in unoperated patients is about 20% (26–28, 31, 33, 34, 101). Surgical annual mortality, including initial surgical mortality, is about 10% (24–28, 30–34, 101, 102). In the VA Cooperative Study (101, 105), patients with left main disease were prospectively randomized and followed for a mean of 22 months. Annual surgical mortality was 9% and medical mortality was 21%, significantly different. Similar results were found in two controlled but nonrandomized studies (28, 34). Hence, at least in the first two years after diagnosis, surgery increases the chance of survival in left main coronary disease.

The long-term prognosis of patients operated upon with two- and three-vessel disease is not clearly improved over the natural history of similar patients not

Table 6 Medical and surgical mortality in controlled study groups

Hospital	Number of surgical patients	Annual Mortaliy (%)		Significance
		Surgical	Medical	
Duke (93)	379	7.5	8.5	NS[a]
Long Beach VA (94, 95)	20	7.5	5	NS[a]
Houston VA (96)	35	4	6	NS[a]

[a] NS = not statistically significant

operated upon. Patients with single-vessel coronary disease have a good long-term prognosis, unoperated approaching less than 2% annual mortality. Those with only the left anterior descending coronary artery involvement have a slightly worse prognosis, 3–4% annual mortality, while those patients with only right coronary or circumflex coronary involvement have a better prognosis (97, 103, 104). However, since the annual mortality with single-vessel disease is so low, one would not expect dramatic improvement in long-term prognosis following surgery. The studies that indicate the results of surgery in patients with single-vessel disease are inconclusive (2, 105) (Table 7). At this time, uncomplicated single-vessel disease is generally not considered an indication for coronary bypass surgery.

In addition to coronary anatomy and ventricular function, the clinical presentation of unstable angina appears to be an important variable in predicting natural survival, especially in certain poor-risk groups (40, 75). Since operative mortality in this group is essentially the same or only slightly higher than in stable angina (35–44), especially after medical stabilization (35, 37–39), and late surgical mortality is the same as in stable angina (35, 36, 38–43, 91), it appears that any difference in medical and surgical treatment depends chiefly on the medical mortality rate. Two randomized studies have reported preliminary data, one showing a surgical advantage (40) and one showing no difference (44). The report showing a surgical advantage was associated with an inordinately high (35%) medical mortality over eight months.

The question of survival will be further clarified over the next several years as data from the VA Cooperative Study and other randomized series are reported.

RESTING VENTRICULAR FUNCTION

Considerable controversy exists over whether bypass surgery improves resting ventricular function. To presume that it does implies that at the time of the relatively elective catheterization, ischemia has progressed to the point of inhibiting contractile function, yet not to the point of causing tissue death, a situation which then could be reversed by revascularization. Disagreement over the reversibility of function can be partly resolved by clarification of (*a*) whether ventricular-segment or overall function is being analyzed, (*b*) the conditions of the postoperative study, and (*c*) the patient grouping.

Table 7 Medical and surgical mortality in patients with left anterior descending coronary artery disease

Study	Number of surgical patients	Annual Mortality (%)		Control group
		Surgery	Control	
Cleveland (2)	Unreported	0	4	same hospital, earlier time period
Alabama (105)	29	3	0	same hospital, same time period

Approximately 50% of patients studied for angina have measurable ventricular dysfunction. Of the preoperatively dyskinetic segments, about 20% will improve their contraction postoperatively (67, 73, 106, 107), although one group reports a much higher rate of improvement (108, 109). Conversely, segment function deteriorates at rates ranging from 15 to 40% (67, 73, 106). Improvement usually occurs before two weeks after surgery and remains at one year, while deterioration occurs both early and late, generally paralleling the time of graft occlusion (106). Factors bearing on segment improvement or deterioration appear to be acuteness of preoperative ischemia, preoperative and perioperative infarction, graft patency, and severity of preoperative hypokinesis (67, 73, 101, 108–112). Overall ventricular function as studied by the ejection fraction is not consistently changed at 1 year (67, 106, 111, 113–116). The ejection fraction usually does not change even in patients where segment function changes, apparently because of compensation by nonischemic portions of the ventricle (106, 111, 117). Increases in the ejection fraction at two weeks after surgery have been reported (108, 109, 112), but the implication that ventricular function is improved has been criticized on the basis of evidence of a high degree of stress and consequent adrenergic stimulation at the time of study (117, 118).

In an attempt to better predict which ventricular segments will improve after surgery, the segmental response of the ventriculogram has been studied after epinephrine (119), nitroglycerin (120, 121), and premature ventricular contractions (122–124). Improvement after these interventions correlates well with improvement after surgery (121, 123, 124).

In summary, surgery causes both improvement and deterioration of segment function and does not consistently affect overall resting ventricular function.

Symptomatic Improvement

The most striking beneficial effect of coronary bypass surgery is improvement of angina. When patients interpret their own change in symptoms, improvement in angina is found in about 85% (range:69–100%) and complete relief in about 70% (range:50–100%) (3, 13–15, 23, 45, 52, 59, 107, 113, 114, 125). This is considerably better than the relief rate of about 25% reported in medically treated patients in the few controlled studies (93, 94, 96). Deterioration in symptoms between the first and second year averages 20% (range:15–40%) of initially improved patients (3, 96, 91). The improvement rates for unstable angina (35–42, 91, 126) and left main disease (28, 30–32), are the same as for stable angina. Failure to improve is associated with poor left ventricular function (127, 128) and graft occlusion (66, 125, 129, 130), although a considerable number of patients with occluded grafts have shown improvement (131, 132). This remarkable symptomatic improvement has also occurred with previous less physiologic procedures to revascularize the heart (133, 134), requiring more objective evaluation of this reported improvement.

About 60% (range:50–84%) of patients are able to perform a significantly greater workload with exercise testing (93, 95, 105, 113, 114, 128) after surgery and about 30% are able to perform a normal workload (128). Likewise, after surgery, 70% reach a higher heart rate before symptoms develop with atrial pacing (96). This rate of improvement is less striking when compared to medical control groups. In one study of 20 surgical patients randomly matched with 20 medical patients, the number of surgically improved patients was initially significantly higher than the medical group, but by 2 years the 76% surgical improvement rate was not statistically different from the 50% medical improvement rate (94, 95). In a nonrandomized study of isolated left anterior descending disease, exercise-test improvement was documented in 10 out of 15 (67%) surgical patients and 7 out of 14 (50%) medical patients (105). In a prospective randomized study of 20 surgical and 21 medical patients, atrial pacing at one year following surgery showed an improvement in 70% of surgical patients and 48% of medical patients, a statistically insignificant difference (96).

As a group, patients can perform about 40% (range:16–89%) more work after surgery (96, 105, 128, 129, 131, 135), and sustain a 25% (range:5–53%) higher heart rate or heart rate–blood pressure product (105, 124, 125, 128, 129, 131, 132, 135, 136). In a prospectively randomized study, work capacity increased 89% in the surgical patients and 42% in the medical patients, a significant difference (136). Objective improvement in exercise capacity correlates with (*a*) graft patency (128, 129, 131, 136), (*b*) good preoperative ventricular function (138), and (*c*) severity of preoperative impairment (128). This last factor is probably self-evident, but bears reemphasis. Patients with angina occurring at low exercise levels improve the most after surgery, while patients with high–exercise threshold angina preoperatively improve very little.

In many cases, surgery, as expected, increases the coronary blood supply, thereby increasing the patient's work capacity. However, in certain instances there is a discrepancy between the expected and observed results. In some patients, there is inadequate revascularization yet improvement in maximum workload and symptoms. One mechanism frequently postulated is that angina is due to a small ischemic segment which then infarcted at surgery. Another interesting observation is that 45% of pain-free patients retain an exercise ECG indicative of ischemia (37, 113, 114, 125), suggesting a disruption of the mechanism where ischemia is interpreted as pain, thereby allowing the patient to exercise past the point of ischemia.

The converse also occurs, where revascularization appears to be adequate, but there is an inappropriate lack of improvement in exercise capacity. In one study, (135), this was shown to be due to a decrease in stroke volume, a change that may be a detrimental effect of surgery, but also is found in deconditioned persons (137).

The most apparent discrepancy in surgical results is the number of patients who report symptomatic improvement, but who show no objective improvement. Several factors account for this improvement in symptoms, including a placebo effect and a less vigorous postoperative life-style. A very important factor is the change in limiting symptom from angina preoperatively (in 65–90% of patients) to fatigue or dyspnea postoperatively (in 80%) (105, 113, 114, 128, 129). Fatigue or dyspnea is a more familiar and desirable limiting symptom to most patients and therefore gives them a sense of improvement even when exercise capacity does not improve. The change in symptoms to fatigue and dyspnea suggests that the limitation to exercise postoperatively is ventricular function and not active ischemia (angina).

UNSTABLE ANGINA

Although the mechanism of unstable angina is not entirely clear, recent evidence suggests that patients with unstable angina develop pain in response to a rise in the heart rate–blood pressure product similar to patients with stable angina, only these rises are spontaneous and of lesser magnitude (138, 139). Propranolol, which lowers the heart rate and blood pressure, is initally used, resulting in disappearance of pain (39). The anatomy (39, 140), operative mortality, and late results of surgery are similar to those found in stable angina; therefore, these patients are medically stabilized and are then generally managed as patients with stable angina. If pain continues despite medical therapy, the patients are then catheterized and considered for bypass surgery.

VARIANT ANGINA PECTORIS

Prinzmetal originally defined variant angina pectoris as a syndrome associated with high proximal obstructions in a single coronary artery, often associated with spasm, with a clinical presentation of angina at rest, nocturnal angina and angina usually not associated with exercise, exertion, or excitment, as occurs in classical angina pectoris (141). Subsequent studies have revealed a less clearly defined group of patients with a variety of coronary artery anatomy, ranging from spasm in normal arteries to severe triple-vessel disease (142–148).

Enthusiasm for surgical revascularization has been strong because of the amenability of isolated proximal lesion in many of the patients to bypass graft surgery, especially following the previously reported grave prognosis associated with variant angina pectoris (146).

Several reports of excellent results of surgical and medical treatment have now been published. However, the management of any given patient must be individualized according to the patient's anatomic disease. Patients with predominant coronary spasm should be managed with drug therapy, while patients with major obstructive lesions should be managed surgically, especially if they do not respond to short-term medical therapy (142, 146–151).

ARRHYTHMIAS

Ventricular tachycardia is an arrhythmia that often has an ischemic origin and results in 40–70% mortality when it is recurrent (152–154). The ventricular tachycardia may be refractory to medical therapy. Several reports of successful suppression of ventricular tachycardia with coronary bypass surgery have appeared (155–157). However, more extensive experiences with larger numbers of patients with medically refractory ventricular tachycardia due to coronary artery disease has lead to some skepticism as to the usefulness of surgery. Since the arrhythmia in several reported patients has not responded to coronary bypass surgery, this suggests that the operative procedure is indicated after all medical therapy, including pacemaker suppression, has been exhausted.

SUMMARY OF SURGICAL THERAPY OF CORONARY ARTERY DISEASE

Patients with single-vessel disease, with normal or mildly abnormal ventricular function (EF >40%), have a good prognosis both for natural survival and long-range symptomatic improvement; therefore medical therapy is strongly recommended. Surgery is considered only if symptoms persist after aggressive medical therapy. It is possible that single-vessel left anterior descending disease is a special variant of this group, and surgery may, with further reports, show an increased survival. However, no adequately designed study has yet suggested this, and it is currently our opinion that patients with single-vessel disease do not have an improved survival following surgery. It is not clear whether surgery improves survival in patients with multivessel disease and normal or mildly abnormal ventricular function (EF >40%). Consequently, cardiologists are divided as to whether to advise surgery in these patients solely for survival. Currently, it is our opinion that these patients should have surgery only for improvement of symptoms after failure of medical therapy. In left main coronary disease the evidence favoring improved survival after surgery has convinced most cardiologists, including ourselves, to recommend surgery. Patients with poor ventricular function (EF <30%) secondary to coronary artery disease often have congestive heart failure and not angina as their chief symptom. Surgery is usually not advisable for these patients, because of the increased operative mortality and lack of improvement in ventricular function. Patients with poor ventricular function with angina are not usually significantly improved by surgery. In patients with moderately abnormal ventricular function (EF=30–40%), relief of angina is frequently obtained, but with some added surgical risk. We recommend surgery in these patients after aggressive medical therapy has failed. Patients with unstable angina are initially medically stabilized, after which they are generally managed as stable angina. Patients with persistence of pain at rest in spite of vigorous medical therapy are usually managed by early catheterization and surgery.

Literature Cited

1. Johnson, W. D., Flemma, R. J., Lepley, D., Ellison, E. H. 1969. Extended treatment of severe coronary artery disease: A total surgical approach. *Ann. Surg.* 170:460–72
2. Sheldon, W. D., Rincon, G., Effler, D. B., Proudfit, W. L., Sones, F. M. Jr. 1973. Vein graft surgery for coronary artery disease: Survival and angiographic results in 1,000 patients. *Circulation* 40:Suppl. III, pp. 184–89
3. McNeer, J. F., Conley, M. J., Starmer, C. F., Behar, V. S., Kong, Y., Peter, R. H., Bartel, A. G., Oldham, H. N. Jr., Young, W. G. Jr., Sabiston, D. C., Rosati, R. A. 1974. Complete and incomplete revascularization at aortocoronary bypass surgery: Experience with 392 consecutive patients. *Am. Heart J.* 88:176–82
4. Wisoff, B. G., Harstein, M. L., Aintablian, A., Hamby, R. I. 1975. Risk of coronary surgery: Two hundred consecutive patients with no hospital deaths. *J. Thorac. Cardiovasc. Surg.* 69:669–73
5. Forker, A. D., Reese, H. E., Weaver, W. F., Wilson, C. S., Buchman, R., Carveth, S. W. 1974. Results of elective aortocoronary saphenous vein graft surgery in a community hospital. *Circulation* 49:334–37
6. Cannom, D. W., Miller, D. C., Shumway, N. E., Fogarty, T. J., Daily, P. O., Hu, M., Brown, B., Harrison, D. C.

1974. The long-term follow-up of patients undergoing saphenous vein bypass surgery. *Circulation* 49:77–85

7. Reul, G. J., Morris, G. C. Jr., Howell, J. F., Crawford, E. S., Stelter, W. J. 1972. Current concepts in coronary artery surgery: A critical analysis of 1,287 patients. *Ann. Thorac. Surg.* 14:243–59
8. Hutchinson, J. E. III, Green, G. E., Mekhjian, H. A., Kemp, H. G. 1974. Coronary bypass grafting in 376 consecutive patients, with three operative deaths. *J. Thorac. Cardiovasc. Surg.* 67:7–16
9. Anderson, R. P., Shahbudin, M. D., Rahimtoola, H., Boncheck, L. I., Starr, A. 1974. The prognosis of patients with coronary artery disease after coronary bypass operations: Time-related progress of 532 patients with disabling angina pectoris. *Circulation* 50:274–82
10. Kouchoukos, N. T., Kirklin, J. W., Oberman, A. 1974. An appraisal of coronary bypass grafting. *Circulation* 50:11–16
11. Moran, J. M., Mulet, J., Caralps, J. M., Pifarre, R. 1974. Coronary revascularization in patients receiving propranolol. *Circulation* 50:Suppl. II, pp. 116–21
12. Hammermeister, K. E., Kennedy, J. W. 1974. Predictors of surgical mortality in patients undergoing direct myocardial revascularization. See Ref. 11, pp. 112–15
13. Najmi, M., Ushiyama, K., Blanco, G., Adam, A., Segal, B. L. 1974. Results of aortocoronary artery saphenous vein bypass surgery for ischemic heart disease. *Am. J. Cardiol.* 33:42–48
14. Hall, R., Dawson, J. T., Cooley, D. A., Hallman, G. L., Wukasch, D. C., Garcia, E. 1976. Coronary artery bypass. *Circulation* 48:Suppl. III, pp. 146–50
15. Collins, J. J. Jr., Cohn, L. H., Sonnenblick, E. H., Herman, M. V., Cohn, P. F., Gorlin, R. 1973. Determinants of survival after coronary artery bypass surgery. See Ref. 14, pp. 132–36
16. Bolooki, H., Sommer, L. S., Ghahramani, A., Cunha, D. 1973. Complications of coronary bypass surgery. See Ref. 14, pp. 120–62
17. Suzuki, A., Kay, C. B., Hardy, J. D. 1973. Direct anastomosis of the bilateral internal mammary artery to the distal coronary artery, without a magnifier, for severe diffuse coronary atherosclerosis. See Ref. 2, pp. 190–97
18. Loop, F. D., Spampinato, N., Siegel, W., Effler, D. B. 1973. Internal mammary artery grafts without optical assistance: Clinical and angiographic analysis of 175 consecutive cases. See Ref. 2, pp. 162–67
19. Cardiac Surgery in the VA. Presented at Meet. Cardiac Surg. Advis. Group, 6th, Washington DC, April, 1975
20. Richter, M. A., Dhurandhar, R. W., O'Meallie, L. P., Rosenberg, M., Glancy, D. L. 1974. Results of all aortocoronary bypass operations in the community hospitals of greater New Orleans. *Circulation* 50:III–67 (Abstr.)
21. Cohn P. F., Gorlin, R., Cohn, L. H., Collins, J. J. Jr. 1974. Left ventricular ejection fraction as a prognostic guide in surgical treatment of coronary and valvular heart disease. *Am. J. Cardiol.* 34:136–41
22. Tyers, G. F. O., Williams, D. R., Pierce, W. S., Babb, J. D. 1974. Results of operative therapy for acquired heart disease: The predictive value of ejection fraction. See Ref. 20, p. 67
23. Oldham, H. N., Kong, Y., Bartel, A. G. 1972. Risk factors in coronary artery bypass surgery. *Arch. Surg.* 105:918–23
24. Alford, W. C. Jr., Shaker, I. J., Thomas, C. S. Jr., Stoney, W. S., Burrus, G. R., Page, H. L. 1974. Aortocoronary bypass in the treatment of left main coronary artery stenosis. *Ann. Thorac. Surg.* 17:247–53
25. Khaja, F., Sharma, S. D., Easley, R. M. Jr., Heinle, R. A., Goldstein, S. 1974. Left main coronary artery lesions: Risks of catheterization; exercise testing and surgery. See Ref. 11, pp. 136–40
26. DeMots, H., Rosch, J., Boncheck, L. I., Anderson, R. P., Starr, A., Rahimtoola, S. H. 1974. Survival in left main coronary artery disease: The role of coronary angiography, coexisting coronary artery disease, and revascularization. *Am. J. Cardiol.* 33:134 (Abstr.)
27. Lim, J. S., Proudfit, W. L., Sones, F. M. Jr. 1975. Left main coronary artery disease: Clinical, arteriographic and hemodynamic appraisal. *Am. J. Cardiol.* 30:791–96
28. Cohen, M. V., Gorlin, R. 1975. Main left coronary artery disease: Clinical experience from 1964–1974. *Circulation* 52:275–85
29. Urschel, H. C. Jr., Razzuk, M. A. 1975. Revascularization of the stenotic left main coronary artery and impaired left ventricle. *J. Thorac. Cardiovasc. Surg.* 69:369–72
30. Zeft, H. M., Manley, J. C., Huston, J. H., Tector, A. J., Auer, J. E., John-

son, W. D. 1974. Left main coronary artery stenosis: Results of coronary bypass surgery. *Circulation* 49:68–76

31. Battock, D. J., Steele, P. P., Davies, H. Left main coronary artery disease—Is surgery always indicated? *Am. J. Cardiol.* 33:125 (Abstr.)
32. Lavine, P., Kimbiris, D., Segal, B. L., Linhart, J. W. Left main coronary artery disease: Clinical, arteriographic and hemodynamic appraisal. *Am. J. Cardiol.* 30:791–96
33. Sung, R. J., Mallon, S. M., Richter, S. E., Ghahramani, A. E., Sommer, L. S., Kaiser, G. A., Myerburg, R. J. 1975. Left main coronary artery obstruction: Follow-up of thirty patients with and without surgery. *Circulation* 52:Suppl. I, pp. 112–18
34. Talano, J. V., Scanlon, P. J., Meadows, W. R., Kahn, M., Pifarre, R., Gunnar, R. M. 1975. Influence of surgery on survival in 145 patients with left main coronary artery disease. See Ref. 33, pp. 105–11
35. Miller, D. C., Cannom, D. S., Fogarty, T. J., Schroeder, J. S., Daily, P. O., Harrison, D. C. 1973. Saphenous vein coronary artery bypass in patients with "preinfarction angina." *Circulation* 47:234–41
36. Conti, C. R., Brawley, R. K., Griffith, L. S. C., Pitt, B., Humphries, J. O., Gott, V. L., Ross, R. S. 1973. Unstable angina pectoris: Morbidity and mortality in 57 consecutive patients evaluated angiographically. *Am. J. Cardiol.* 32:745–50
37. Berndt, T. B., Miller, D. C., Silverman, J. F., Stinson, E. B., Harrison, D. C., Schroeder, J. S. 1975. Coronary bypass surgery for unstable angina pectoris: Clinical follow-up and results of postoperative treadmill electrocardiograms. *Am. J. Med.* 58:171–76
38. Kouchoukos, N. T., Russel, R. O. Jr., Moraski, R. E., Karp, R. B., Oberman, A., Rackley, C. E. 1975. Surgical treatment of unstable angina pectoris: Results in 65 patients. *Am. J. Cardiol.* 35:149 (Abstr.)
39. Fischl, S. J., Herman, M. V., Gorlin, R. 1973. The intermediate coronary syndrome: Clinical, angiographic and therapeutic aspects. *N. Engl. J. Med.* 288:1193–98
40. Bertolasi, C. A., Tronge, J. E., Carreno, C. A., Jalon, J., Vega, M. R. 1974. Unstable angina—prospective and randomized study of its evaluation, with and without surgery: Preliminary report. *Am. J. Cardiol.* 33:201–8
41. Matloff, J. M., Sustaita, H., Chatterjee, K., Chaux, A., Marcus, H. S., Swan, H. J. C. 1975. The rationale for surgery in preinfarction angina. *J. Thorac. Cardiovasc. Surg.* 69:73–81
42. Bolooki, H., Sommer, L., Kaiser, G. A., Vargas, A., Ghahramani, A. 1974. Long-term follow-up in patients receiving emergency revascularization for intermediate coronary syndrome. *J. Thorac. Cardiovasc. Surg.* 68:90–100
43. Hammond, G. L., Poirier, R. A. 1975. Surgical management for acute coronary insufficiency with three years' follow-up. *J. Thorac. Cardiovasc. Surg.* 69:625–30
44. Conti, C. R., Gilbert, J. B., Hodges, M., Hutter, A. M. Jr., Kaplan, E. M., Newell, J. B., Resnekov, L., Rosati, R. A., Ross, R. S., Russell, R. O. Jr., Schroeder, J. S., Wolk, M. 1975. Unstable angina pectoris: Randomized study of surgical vs. medical therapy. *Am. J. Cardiol.* 35:129 (Abstr.)
45. Morris, G. C. Jr., Reul, G. J., Howell, J. F. 1972. Follow-up results of distal coronary artery bypass for ischemic heart disease. *Am. J. Cardiol.* 29: 180–85
46. Ashor, G. W., Meyer, B. W., Lindesmith, G. G., Stiles, Q. R., Walker, G. H., Tucker, B. L. 1973. Coronary artery disease: Surgery in 100 patients 65 years of age and older. *Arch. Surg.* 107:30–33
47. Anderson, R. P., Hodam, R., Wood, J., Starr, A. 1972. Direct revascularization of the heart: Early clinical experience with 200 patients. *J. Thorac. Cardiovasc. Surg.* 63:353–59
48. Verska, J. J., Walker, W. J. 1975. Aortocoronary bypass in the diabetic patient. *Am. J. Cardiol.* 35:744–77
49. Engelman, R. M., Bhat, J. G., Glassman, E., Spencer, F. C., Boyd, A. D., Pasternak, B. S., Reed, G. E., Isom, O. W. The influence of diabetes and hypertension on the results of coronary revascularization. *Am. J. Cardiol.* 35:135 (Abstr.)
50. Keon, W. J., Bedard, P., Shankar, K. R., Akyurekli, Y., Nino, A., Berkman, F. 1973. Experience with emergency aortocoronary bypass grafts in the presence of acute myocardial infarction. See Ref. 14, pp. 151–55
51. Hill, J. D., Kerth, W. J., Kelly, J. J., Selzer, A., Armstrong, W., Popper, R. W., Langston, M. F., Cohn, K. E. 1971. Emergency aortocoronary bypass for

impending or extending myocardial infarction. *Circulation* 43:Suppl. I, pp. 105–10

52. Leinbach, R. C., Gold, H. K., Dinsmore, R. E., Mundth, E. D., Buckley, M. J., Austen, W. G., Sanders, C. A. 1973. The role of angiography in cardiogenic shock. See Ref. 14, pp. 95–98
53. Dawson, J. T., Hall, R. J., Hallman, G. L., Cooley, D. A. 1974. Mortality in patients undergoing coronary artery bypass surgery after myocardial infarction. *Am. J. Cardiol.* 33:483–86
54. Cheanvechai, C., Effler, D. B., Loop, F. D., Groves, L. K., Sheldon, W. C., Razavi, M., Sones, F. M. 1973. Emergency myocardial revascularization. *Am. J. Cardiol.* 32:901–8
55. Mundth, E. D., Yurchak, P. M., Buckley, M. J., Leinbach, R. C., Kantrowitz, A., Austen, W. G. 1970. Circulatory assistance and emergency direct coronary-artery surgery for shock complicating acute myocardial infarction. *N. Engl. J. Med.* 283:1382–84
56. Willerson, J. T., Curry, G. C., Watson, J. T., Leshin, S. J., Ecker, R. R., Platt, M. R., Mullins, C. B., Sugg, W. L. 1975. Intraaortic balloon counterpulsation in patients in cardiogenic shock, medically refractory left ventricular failure and/or recurrent ventricular tachycardia. *Am. J. Med.* 58:183–91
57. Cohn, L. H., Fogarty, T. J., Daily, P. O., Shumway, N. E. 1971. Emergency coronary artery bypass. *Surgery* 70: 821–29
58. Cheanvechai, C., Effler, D. B., Groves, L. K. 1974. Triple bypass graft for the treatment of severe triple coronary vessel disease. *Ann. Thorac. Surg.* 17: 545–54
59. Lawrence, G. H., Riggins, R. C. K., Hipp, R., Johnston, R. R. 1973. Status of one hundred patients after coronary artery bypass surgery. *Am. J. Surg.* 126:277–85
60. Brewer, D. L., Bilbro, R. H., Bartel, A. G. 1973. Myocardial infarction as a complication of coronary bypass surgery. *Circulation* 47:58–64
61. Schrank, J. P., Slabaugh, T. K., Beckwith, J. R. The incidence and clinical significance of ECG-VCG changes of myocardial infarction following aortocoronary saphenous vein bypass surgery. *Am. Heart J.* 87:46–54
62. Assad-Morell, J. L., Frye, R. L., Connolly, D. C. 1975. Relation of intraoperative or early postoperative transmural myocardial infarction to patency of aortocoronary bypass grafts and to diseased ungrafted coronary arteries. *Am. J. Cardiol.* 35:767–73
63. Dixon, S. H. Jr., Limbird, L. E., Roe, C. R., Wagner, G. S., Oldham, H. N. Jr., Sabiston, D. C. Jr. 1973. Recognition of postoperative acute myocardial infarction: Application of isoenzyme techniques. See Ref. 14, pp. 137–40
64. Alderman, E. L., Matlof, H. J., Shumway, N. E., Harrison, D. C. 1973. Evaluation of enzyme testing for the detection of myocardial infarction following direct coronary surgery. *Circulation* 48:135–40
65. Rose, M. R., Glassman, E., Isom, O. W., Spencer, F. C. 1974. Electrocardiographic and serum enzyme changes of myocardial infarction after coronary artery bypass surgery. *Am. J. Cardiol.* 33:215–20
66. Alderman, E. L., Matlof, H. J., Wexler, L., Shumway, N. E., Harrison, D. C. 1973. Result of direct coronary-artery surgery for the treatment of angina pectoris. *N. Engl. J. Med.* 288:535–39
67. Shepherd, R. L., Itscoitz, S. B., Glancy, D. L. 1975. Deterioration of myocardial function following aortocoronary bypass operation. *Circulation* 49:467–75
68. Vlodaver, Z., Edwards, J. E. 1973. Pathologic analysis in fatal cases following saphenous vein coronary arterial bypass. *Chest* 64:555–63
69. Bonchek, L. I., Rahimotoola, S. H., Chaitman, B. R., Rosch, J., Anderson, R. P., Starr, A. 1974. Vein graft occlusion: Immediate and late consequences and therapeutic implications. See Ref. 11, pp. 84–97
70. Maurer, B. J., Oberman, A., Holt, J. H. Jr., Kouchoukos, N. T., Jones, W. B., Russell, R. O. Jr., Reeves, T. J. 1974. Changes in grafted and nongrafted coronary arteries following saphenous vein bypass grafting. *Circulation* 50: 293–300
71. Malinow, M. R., Kremkau, E. L., Kloster, F. E., Bonchek, L. I., Rosch, J. 1973. Occlusion of coronary arteries after vein bypass. *Circulation* 47:1211–16
72. Glassman, E., Spencer, F. C., Krauss, K. R., Weisinger, B., Isom, O. W. 1974. Changes in the underlying coronary circulation secondary to bypass grafting. See Ref. 11, pp. 84–97
73. Griffith, L. S. C., Achuff, S. C., Conti, C. R., Humphries, O., Brawley, R. K., Gott, V. L., Ross, R. S. 1972. Changes in intrinsic coronary circulation and segmental ventricular contractility after

saphenous vein coronary bypass graft surgery. *Trans. Assoc. Am. Physicians* 85:248–58

74. Bourassa, M. G., Goulet, C., Lesperance, J. 1973. Progression of coronary arterial disease after aortocoronary bypass grafts. See Ref. 14, pp. 127–31
75. Bousvaros, G., Piracha, A. R., Chaudhry, M. A., Grant, C., Older, T. M., Pifarre, R. 1972. Increase in severity of proximal coronary disease after successful distal aortocoronary grafts. *Circulation* 46:870–79
76. Levine, J. A., Bechtel, D. J., Gorlin, R., Cohn, P. F., Herman, M. V., Cohn, L. H., Collins, J. J. Jr. 1975. Coronary artery anatomy before and after direct revascularization surgery: Clinical and cinearteriographic studies in 67 selected patients. *Am. Heart J.* 89:561–70
77. Pasternak, R., Cohn, K., Selzer, A., Langston, M. F. Jr. 1975. Enhanced rate of progression of coronary artery disease following aortocoronary saphenous vein bypass surgery. *Am. J. Med.* 58:166–70
78. Grondin, C. M., Lesperance, J., Bourassa, M. G., Pasternac, A., Campeau, L., Grondin, P. 1974. Serial angiographic evaluation in 60 consecutive patients with aortocoronary artery vein grafts 2 weeks, 1 year, and 3 years after operation. *J. Thorac. Cardiovasc. Surg.* 67:1–5
79. Walker, J. A., Friedberg, H. D., Flemma, R. J., Johnson, W. D. 1972. Determinants of angiographic patency of aortocoronary vein bypass grafts. *Circulation* 45:Suppl. I, pp. 86–90
80. Itscoitz, S. B., Redwood, D. R., Grauer, L. E., Reis, R. L., Stinson, E. B., Epstein, S. E. 1974. Long-term durability of patent saphenous vein aortocoronary bypass grafts. *Am. J. Cardiol.* 33:146 (Abstr.)
81. Unni, K. K., Kottke, B. A., Titus, J. L., Frye, R. L., Wallace, R. B., Brown, A. L. 1974. Pathologic changes in aortocoronary saphenous vein grafts. *Am. J. Cardiol.* 34:526–32
82. Kennedy, J. H., Wieting, D. W., Hwang, N. H. C., Anderson, M. S., Bayardo, R. J., Howell, J. F., DeBakey, M. E. 1974. Hydraulic and morphologic study of fibrous intimal hyperplasia in autogenous saphenous vein bypass grafts. *J. Thorac. Cardiovasc. Surg.* 67:805–13
83. Jones, M., Conkle, D. M. Ferrans, V. J., Roberts, W. C., Levine, F. H., Melvin, D. B., Stinson, E. B. 1973. Lesions observed in arterial autogenous vein grafts. Light and electron microscopic evaluation. See Ref. 14, pp. 198–210
84. Grondin, C. M., Meere, C., Castonguay, Y., Lepage, G., Grondin, P. 1971. Progressive and late obstruction of an aortocoronary venous bypass graft. *Circulation* 43:698–702
85. Barboriak, J. J., Pintal, K., Korns, M. E. 1974. Atherosclerosis in aortocoronary vein grafts. *Lancet* 27:621
86. Lesperance, J., Bourassa, M. G., Biron, P., Campeau, L., Saltiel, J. 1972. Aorta to coronary artery saphenous vein grafts: Preoperative angiographic criteria for successful surgery. *Am. J. Cardiol.* 30:459–65
87. Rosch, J., Dotter, C. T., Antonovic, R., Bonchek, L., Starr. A. 1973. Angiographic appraisal of distal vessel suitability for aortocoronary bypass graft surgery. *Circulation* 48:202–12
88. Grondin, C. M., Lepage, G., Castonguay, Y. R., Meere, C., Grondin, P. 1971. Aortocoronary bypass graft: Initial blood flow through the graft, and early postoperative patency. *Circulation* 44:815–19
89. Dunkman, W. B., Perloff, J. K., Kastor, J. A., Shelburne, J. C. 1974. Medical perspectives in coronary artery surgery —A caveat. *Ann. Intern. Med.* 81: 817–37
90. Mundth, E. D., Austen, W. G. 1975. Surgical measures for coronary heart disease. *N. Engl. J. Med.* 293:13–19, 75–80, 124–30
91. Tecklenberg, P. L., Alderman, E. L., Miller, D. C., Shumway, N. E., Harrison, D. C. Changes in survival and symptom relief in a longitudinal study of patients after bypass surgery. See Ref. 33, pp. 98–104
92. Cooley, D. A., Dawson, J. T., Hallman, G. L., Sandiford, F. M. 1973. Aortocoronary saphenous vein bypass: Results in 1,492 patients with particular reference to patients with complicating features. *Ann. Thorac. Surg.* 16:380–90
93. McNeer, J. F., Starmer, C. F., Bartel, A. G., Behar, V. S., Kong, Y., Peter, R. H., Rosati, R. A. 1974. The nature of treatment selection in coronary artery disease: Experience with medical and surgical treatment of a chronic disease. *Circulation* 49:606–14
94. Aronow, W. S., Stemmer, E. A. 1975. Two-year follow-up of angina pectoris: Medical or surgical therapy. *Ann. Intern. Med.* 82:208–17

95. Aronow, W. S., Stemmer, E. A. 1974. Bypass graft surgery versus medical therapy of angina pectoris. *Am. J. Cardiol.* 33:415–20
96. Mathur, V. S., Guinn, G. A., Anastassiades, L. C., Chahine, R. A., Korompai, F. L., Montero, A. C., Luchi, R. J. 1975. Surgical treatment for stable angina pectoris: Prospective randomized study. *N. Engl. J. Med.* 292:709–10
97. Reeves, T. J., Oberman, A., Jones, W. B., Sheffield, L. T. 1974. Natural history of angina pectoris. *Am. J. Cardiol.* 33:423–30
98. Webster, J. S., Moberg, C., Rincon, G. 1974. Natural history of severe proximal coronary artery disease as documented by coronary cineangiography. *Am. J. Cardiol.* 33:195–200
99. Nelson, G. R., Cohn, P. F., Gorlin, P. 1975. Prognosis in medically-treated coronary artery disease: Influence of ejection fraction compared to other parameters. *Circulation* 52:408–12
100. Oberman, A., Jones, W. B., Riley, C. P., Reeves, T. J., Sheffield, L. T., Turner, M. E. 1972. Natural history of coronary artery disease. *Bull. NY Acad. Med.* 48:1109–15
101. Takaro, T., Hultgren, H. N., Detre, K. M. 1975. VA Cooperative Study of coronary arterial surgery: II. Left main disease. See Ref. 33, p. 143
102. Scanlon, P. J., Talano, J. V., Moran, J. F., Gunnar, R. M. 1974. Main left coronary artery disease. *Am. J. Cardiol.* 33:169 (Abstr.)
103. Burggraf, G. W., Parker, J. O. 1975. Prognosis in coronary artery disease: Angiographic, hemodynamic, and clinical factors. *Circulation* 51:146–56
104. Bruschke, A. V. G., Proudfit, W. L., Sones, F. M. Jr. 1973. Progress study of 590 consecutive nonsurgical cases of coronary disease followed 5–9 years. I. Arteriographic correlations. *Circulation* 47:1147–53
105. Kouchoukos, N. T., Oberman, A., Russell, R. O. Jr., Jones, W. B. 1975. Surgical versus medical treatment of occlusive disease confined to the left anterior descending coronary artery. *Am. J. Cardiol.* 35:836–42
106. Bourassa, M. G., Lesperance, J., Campeau, L., Saltiel, J. 1972. Fate of left ventricular contraction following aortocoronary venous grafts: Early and late postoperative modifications. *Circulation* 46:724–30
107. Griffith, L. S. C., Achuff, S. C., Conti, C. R., Humphries, J. O., Brawley, R. K., Gott, V. L., Ross, R. S. 1973. Changes in intrinsic coronary circulation and segmental ventricular motion after saphenous-vein coronary bypass graft surgery. *N. Engl. J. Med.* 288:589–94
108. Chatterjee, K., Swan, H. J. C., Parmley, W. W., Sustaita, H., Marcus, H., Matloff, J. 1972. Depression of left ventricular function due to acute myocardial ischemia and its reversal after aortocoronary saphenous-vein bypass. *N. Engl. J. Med.* 286:1117–22
109. Chatterjee, K., Swan, H. J. C., Parmley, W. W., Sustaita, H., Marcus, H. S., Matloff, J. 1973. Influence of direct myocardial revascularization on left ventricular asynergy and function in patients with coronary heart disease: With and without previous myocardial infarction. *Circulation* 47:276–86
110. Rees, G., Bristow, J. D., Kremkau, E. L., Green, G. S., Herr, R. H., Griswold, H. E., Starr, A. 1971. Influence of aortocoronary bypass surgery on left ventricular performance. *N. Engl. J. Med.* 284:1116–20
111. Arbogast, R., Solignac, A., Bourassa, M. G. 1973. Influence of aortocoronary saphenous vein bypass surgery on left ventricular volumes and ejection fraction: Comparison before and one year after surgery in 51 patients. *Am. J. Med.* 54:290–96
112. Hamby, R. I., Tabrah, F., Aintablian, A., Hartstein, M. L., Wisoff, B. G. 1974. Left ventricular hemodynamics and contractile pattern after aortocoronary bypass surgery: Factors affecting reversibility of abnormal left ventricular function. *Am. Heart J.* 88:149–59
113. Conti, C. R., Page, E. E., Humphries, J. O., Pitt, B., Ross, R. S. 1972. Objective evaluation of aortico-coronary vein bypass surgery. *Trans. Assoc. Am. Physicians* 85:272–79
114. Hammermeister, K. E., Kennedy, J. W., Hamilton, G. W., Stewart, D. K., Gould, K. L., Lipscomb, K., Murray, J. A. 1974. Aortocoronary saphenous-vein bypass: Failure of successful grafting to improve resting left ventricular function in chronic angina. *N. Engl. J. Med.* 290:186–92
115. Levine, J. A., Bechtel, D. J., Cohn, P. F., Herman, M. V., Gorlin, R., Cohn, L. H., Collins, J. J. Jr. 1975. Ventricular function before and after direct revascularization surgery: A proposal for an index of vascularization to correlate

angiographic and ventriculographic findings. *Circulation* 51:1071–78

116. Morch, J., Morton, B., McLaughlin, P. 1974. Late results of aortocoronary bypass grafts in 100 patients with stable angina pectoris. *Can. Med. Assoc. J.* 111:529–32
117. Bourassa, M. G. 1973. Left ventricular performance following direct myocardial revascularization (editorial). *Circulation* 48:915–16
118. Hammermeister, K. E., Kennedy, J. W. 1973. Myocardial revascularization and ventricular performance (editorial). *Circulation* 48:450–51
119. Horn, H. R., Teichholz, L. E., Cohn, P. F., Herman, M. V., Gorlin, R. 1974. Augmentation of left ventricular contraction pattern in coronary artery disease by an inotropic catecholamine: The epinephrine ventriculogram. *Circulation* 49:1063–71
120. Banka, V. S., Bodenheimer, M. M., Helfant, R. H. 1974. Determinants of reversible asynergy: Effect of pathologic Q waves, coronary collaterals, and anatomic location. *Circulation* 50:714–19
121. Helfant, R. H., Pine, R., Meister, S. G., Feldman, M. S., Trout, R. G., Banka, V. S. Nitroglycerin to unmask reversible asynergy: Correlation with post coronary bypass ventriculography. *Circulation* 108–13
122. Dyke, S. H., Cohn, P. F., Gorlin, R., Sonnenblick, E. H. 1974. Detection of residual myocardial function in coronary artery disease using post-extra systolic potentiation. *Circulation* 50: 694–99
123. Cohn, P. F., Gorlin, R., Herman, M. V., Sonnenblick, E. H., Horn, H. R., Cohn, L. H., Collins, J. J. Jr. 1975. Relation between contractile reserve and prognosis in patients with coronary artery disease and a depressed ejection fraction. *Circulation* 51:414–20
124. Hamby, R. I., Aintablian, A., Wisoff, B. G., Hartstein, M. L. 1975. Response of the left ventricle in coronary artery disease to postextrasystolic potentiation. *Circulation* 51:428–35
125. Bartel, A. G., Behar, V. S., Peter, R. H., Orgain, E. S., Kong, Y. 1973. Exercise stress testing in evaluation of aortocoronary bypass surgery: Report of 123 patients. *Circulation* 48:141–48
126. Weintraub, R. M., Voukydis, P. C., Aroesty, J. M. 1974. Treatment of preinfarction angina with intraaortic balloon counterpulsation and surgery. *Am. J. Cardiol.* 34:809–14
127. Yatteau, R. F., Peter, R. H., Behar, V. A., Bartel, A. G., Rosati, R. A., Kong, Y. 1974. Ischemic cardiomyopathy: The myopathy of coronary artery disease. Natural history and results of medical versus surgical treatment. *Am. J. Cardiol.* 34:520–25
128. Lapin, E. S., Murray, J. A., Bruce, R. A., Winterscheid, L. 1973. Changes in maximal exercise performance in the evaluation of saphenous vein bypass surgery. *Circulation* 47:1164–73
129. Balcon, R., Honey, M., Rickards, A. F., Sturridge, M. F., Walsh, W., Wilkinson, R. K., Wright, J. E. C. 1974. Evaluation by exercise testing and atrial pacing of results of aortocoronary bypass surgery. *Br. Heart J.* 36:841–53
130. Surgical Research Society. 1973. Meeting held London, 1973. *Br. J. Surg.* 60:305–23
131. Dodek, A., Kassebaum, D. G., Griswold, H. E. 1973. Stress electrocardiography in the evaluation of aortocoronary bypass surgery. *Am. Heart J.* 86:292–307
132. DiLuzio, V., Roy, P. R., Sowton, E. 1974. Angina in patients with occluded aorto-coronary vein grafts. *Br. Heart J.* 36:139–46
133. Beecher, H. K. 1961. Surgery as placebo: A quantitative study of bias. *J. Am. Med. Assoc.* 176:1102–7
134. Vansant, J. H., Muller, W. H. Jr. 1960. Surgical procedures to revascularize the heart: A review of the literature. *Am. J. Surg.* 100:572–83
135. McDonough, J. R., Danielson, R. A., Wills, R. E., Vine, D. L. 1974. Maximal cardiac output during exercise in patients with coronary artery disease. *Am. J. Cardiol.* 33:23–29
136. Mathur, V. S., Guinn, G. A. 1975. Prospective randomized study of coronary bypass surgery in stable angina: The first 100 patients. See Ref. 33, pp. 133–40
137. Astrand, P. O., Rodahl, K. 1970. *Textbook of Work Physiology.* New York: McGraw-Hill
138. Cannom, D. S., Harrison, D. C., Schroeder, J. S. 1974. Hemodynamic observations in patients with unstable angina pectoris. *Am. J. Cardiol.* 33:17–22
139. Littler, W. A., Honour, A. J., Sleight, P., Stott, F. D. 1973. Direct arterial pressure and the electrocardiogram in unrestricted patients with angina pectoris. *Circulation* 48:125–34
140. Guthrie, R. B., Vlodaver, Z., Nicoloff, D. M., Edwards, J. E. 1975. Pathology

of stable and unstable angina pectoris. *Circulation* 51:1059–63
141. Prinzmetal, M., Kennamer, R., Merliss, R., Wada, T., Bor, N. 1959. Angina pectoris. I. A variant form of angina pectoris; preliminary report. *Am. J. Med.* 27:375–88
142. Donsky, M. S., Harris, M. D., Burry, G. C., Blomqvist, C. G., Willerson, J. T., Mullins, C. B. 1975. Variant angina pectoris: a clinical and coronary arteriographic spectrum. *Am. Heart J.* 89:571–78
143. Demany, M. A., Tambe, A., Zimmerman, H. A. 1968. Coronary artery spasm. *Dis. Chest* 53:714–21
144. Gianelly, R., Mugler, F., Harrison, D. C. 1968. Prinzmetal's variant of angina pectoris with only slight coronary atherosclerosis. *Calif. Med.* 108:129–36
145. Whiting, R. B., Klein, M. D., Vander Veer, J., Lown, B. 1970. Variant angina pectoris. *N. Engl. J. Med.* 282:709–12
146. Silverman, M. E., Flamm, M. D. Jr. 1971. Variant angina pectoris: anatomic findings and prognostic implications. *Ann. Intern. Med.* 75:339–44
147. MacAlpin, R. M., Kattus, A. A., Alvaro, A. B. 1973. Angina pectoris at rest with preservation of exercise capacity: Prinzmetal's variant angina. *Circulation* 47:946–58
148. Fernandez, D., Rosenthal, J. E., Cohen, L. S., Hammond, G., Wolfson, S. 1973. Alcohol-induced Prinzmetal variant angina. *Am. J. Cardiol.* 32:238–39
149. Betriu, A., Solignac, A., Bourassa, M. G. 1974. The variant form of angina: diagnostic and therapeutic implications. *Am. Heart J.* 87:272–78
150. Hart, N. J., Silverman, M. E., King, S. B. 1974. Variant angina pectoris caused by coronary artery spasm. *Am. J. Med.* 56:269–74
151. Epstein, S. E., Redwood, D. R., Goldstein, R. E., Beiser, G. D., Rosing, D. R., Glancy, D. L., Reis, R. L., Stinson, E. B. 1971. Angina pectoris: pathophysiology, evaluation, and treatment. *Ann. Intern. Med.* 75:263–96
152. Armbrust, C. A. Jr., Levine, S. A. 1950. Paroxysmal ventricular tachycardia: A study of one hundred and seven cases. *Circulation* 1:28–33
153. Day, H. W. 1968. Progress in cardiology: Acute coronary care; a five year report. *Am. J. Cardiol.* 21:252–59
154. Ecker, R. R., Mullins, C. B., Grammer, J. C., Rea, W. J., Atkins, J. M. Control of intractable ventricular tachycardia by coronary revascularization. *Circulation* 44:666–73
155. Lambert, C. J., Adam, M., Geisler, G. F., Verzosa, E., Naxarian, M., Mitchel, B. F. 1971. Emergency myocardial revascularization for impending infarctions and arrhythmias. *J. Thorac. Cardiovasc. Surg.* 62:522–28
156. Nakhjavan, F. K., Morse, D. P., Nichols, H. T., Goldberg, H. 1971. Emergency aortocoronary bypass. Treatment of ventricular tachycardia due to ischemic heart disease. *J. Am. Med. Assoc.* 216:2138–40
157. Ecker, R. R., Mullins, C. B., Grammer, J. C., Rea, W. J., Atkins, J. M. 1971. Control of intractable ventricular tachycardia by coronary revascularization. *Circulation* 44:666–70

Ann. Rev. Med. 1977. 28:291–98
Copyright © 1977 by Annual Reviews Inc. All rights reserved

COLD AGGLUTININS AND COLD-AGGLUTININ DISEASE

◆7243

Michael M. Frank, M.D., John P. Atkinson, M.D., and James Gadek, M.D.

Laboratory of Clinical Investigation, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland 20014

Since the original descriptions of cold-reactive antibodies at the turn of the century, this group of antibodies and the diseases associated with their appearance have become the focus of a great deal of interest (1–5). Early investigators recognized that a group of these patients had paroxysmal hemoglobinuria on exposure to cold, and a subgroup of patients with a cold-reacting, hemolytic antibody (Donath-Landsteiner) was identified (6). In many of these patients the disease was associated with long-standing syphilitic infection. More recent investigations have centered on cold agglutinins, cold-reactive antierythrocyte antibodies present in several clinical settings, which are directed at defined groups of erythrocyte antigens and which usually are weakly lytic in vivo (1–5, 7). These antibodies are usually found in patients with cold-agglutinin disease or may occur transiently following a number of acute infectious illnesses, the classic example being Mycoplasma pneumonia (8, 9).

Two main areas of recent investigation have been particularly fruitful. The first of these concerns the immunochemical characterization of the reaction; that is, the chemical definition of the erythrocyte antigens involved, the chemical definition of the antibodies, and studies to determine why this antigen-antibody complex agglutinates at temperatures below 37°C. The second group of studies has examined the pathophysiologic basis of cold-agglutinin disease and the events that occur on the erythrocyte surface following its interaction with cold agglutinin, leading to its destruction. The latter area of investigation has been of particular concern to this laboratory, and we focus on this problem in this review. We conclude by summarizing the limited information available on the etiology of the cold agglutinins in this interesting and perplexing group of diseases.

Cold agglutinins are defined as antibodies capable of agglutinating erythrocytes at temperatures below 37°C. Unlike most agglutinating antibodies, cold agglutinins require this lowered temperature to manifest their effects; they do not agglutinate cells at 37°C, and in fact dissociate from their site of binding as the temperature of the reaction mixture is raised to 37°C (1–5, 10–12). Very low titers of cold agglutinins are found in many normal individuals. Elevated levels are usually identified in two general clinical settings (1, 3). Most commonly, they appear transiently following infectious illnesses—often Mycoplasma pneumonia (8, 9). Only rarely is their appearance in this setting associated with clinical hemolysis. Less frequently, cold agglutinins occur as monoclonal antibodies, often in the absence of identifiable underlying disease (1–7). Cold-agglutinin disease is often associated with mild hemolytic anemia, and it is occasionally punctuated by severe hemolytic crises, usually following ambient exposure to cold (1, 3). This disease tends to occur in the elderly, with a peak incidence reported at ages 50–60 years (3). Although these two general clinical settings in which cold agglutinins are found have certain features in common, the most complete studies have usually examined the antibodies found in the long-standing cold-agglutinin disease.

The majority of patients with this syndrome have in their circulation an IgM antibody with specificity for the Ii group of erythrocyte antigens. There has been a great deal of recent work concerned with the characterization of these antigens. Biochemically, this group of cell-membrane antigens appears to be related to the precursors of the ABO-Lewis blood group of glycoproteins (4, 5, 7). As originally defined, the specificity of IgM cold agglutinin was toward the I antigen, an antigen present on the erythrocytes of more than 95% of adults, but functionally present in smaller quantities on erythrocyte membranes of newborns. Thus the antibody present in most of these patients would agglutinate the erythrocytes of most adults at temperatures below 37°C, but would not agglutinate cord-blood erythrocytes. This contrasts with the antibodies present in many patients following infectious illnesses, which react with an antigen, termed i, that is present on cord-blood erythrocytes. This antigen (i) is also present on the erythrocytes of adults, although the effective antigenic density is lower than that found on the erythrocytes of newborns (13).

Cold agglutinins found in cold-agglutinin disease have many of the characteristics of a myeloma protein (14). They have highly restricted heterogeneity, and the protein in an individual patient may appear to have a single electrophoretic mobility and a single amino-acid sequence (4, 5, 7). Under normal circumstances, antibodies found after antigenic challenge are quite heterogeneous, with a wide range of electrophoretic mobilities and amino-acid sequences. In keeping with the restricted heterogeneity of cold agglutinins is the fact that the vast majority have only one of the two light-chain types found in normal antibodies. Most cold agglutinins have kappa light-chain specificity only. However, a comparison of the cold agglutinin found in the serum of one individual to that found in another reveals considerable differences. They may have different electrophoretic mobility and chemical structure. There may be a considerable difference in the temperature range over which they show maximal erythrocyte binding, i.e. they are said to differ in thermal amplitude. This is clinically important. Antibodies with high thermal amplitude, that is, antibodies that react with the erythrocyte membrane at temperatures nearer physiologic are more likely to cause clinical illness than those antibodies that only interact with erythrocytes at very low temperatures (1, 11). Different cold agglutinins may show differences in affinity for their antigen and differences in the efficiency with which they mediate some of the biological effects of antibody, for example, complement activation. The antigenic specificity of the various antibodies has also been shown to differ widely. One way of looking at the antigenic specificity of these antibodies is that they may all recognize some aspect of the highly complex I antigen, but all recognize different parts of the same antigenic structure (15). Recent work has divided the anti-I antibodies into six groupings that reflect these different specificities on the basis of inhibition of precipitation with purified blood-group substances (16). It is of interest that these antibodies appear to interact with partially purified I antigens extracted from erythrocyte surfaces equally well at 37°C and 0°C (17). The current feeling is that there is no major change in the structure of the antibody or the antibody-combining site at lowered temperatures, but that the reason that cold agglutinins function only, or with greater efficiency, in the cold is related to alterations in the structure or availability of the glycosidic moiety in the erythrocyte membrane at lowered temperatures. Thus, both primary antibody effects such as binding to antigen and secondary effects such as complement fixation occur with cold agglutinins at 37°C when the I antigen is extracted from the erythrocyte membrane.

Some years ago it was shown that rabbits could be immunized to cold agglutinins, and the resulting antisera absorbed with various non–cold agglutinin IgM proteins in order to make antibody specific for the immunizing cold agglutinin (18). Such rabbit antibody was shown to have specificity for antigens in the variable region of the cold-agglutinin IgM molecule. Since the rabbit antibody reacted preferentially with the immunizing antigen and not with other non–cold-agglutinin IgM proteins, it was said to display idiotypic specificity. Usually such idiotypic antibodies, with specificity for the variable region of myeloma proteins, show very little cross-reactivity with other myeloma proteins; however, such idiotypic antibodies were shown to demonstrate cross-reactivity between several cold agglutinins (18). Recently, this work has been broadened and extensive idiotypic cross-reactivity has been demonstrated among many cold agglutinins (19). One interpretation of these findings is that there is considerable similarity in the chemical structure of the variable region of anti-I cold agglutinin from different individuals. Thus, there may be considerable homology in the binding sites of the cold-agglutinin antibodies made by different individuals with cold-agglutinin disease.

Although this brief review has focused on anti-I IgM antibodies directed to erythrocyte antigens, a number of other species of antibody with cold agglutinin specificity have been described. Not all cold agglutinins are directed at the Ii group of antigens (5). Another series of antigens on the erythrocyte surface, the Pr group of antigens, has been identified (20). These antigens are termed Pr because unlike the I antigens, which resist destruction by proteases such as papain and ficin, the Pr group of antigens are protease-sensitive. These antigens are also thought to be glycosidic residues on the erythrocyte surface. They are present on both cord and adult red cells. Antibodies with specificity for this antigen are commonly found after infectious diseases, particularly infectious mononucleosis, and these antibodies occur fairly frequently in the IgA group as well as IgM. IgG antibodies of this class have been identified as well. It is also of interest that a number of these antigens, particularly those of the Ii series, are present not only on erythrocytes, but also on B- and T-lymphocytes, granulocytes, and cells of the macrophage series (21, 22). Thus, defects in the function of cells other than those in the erythrocyte series might be expected to be present in this disease and are worthy of investigation.

Considerable progress has also been made in elucidating the pathophysiology of erythrocyte destruction in cold-agglutinin disease. It has long been known that erythrocytes from patients with cold-agglutinin disease are Coombs-positive and contain membrane-bound complement components on their surfaces (10). It has also been shown that if these Coombs-positive erythrocytes are taken from the patient, radiolabeled, and reinfused into the erythrocyte donor, they have a more prolonged survival in the donor than do normal erythrocytes (23, 24). Moreover, this same phenomenon can be duplicated in vitro (25). Coombs-positive erythrocytes from patients with cold-agglutinin disease are less susceptible to in vitro hemolysis mediated by cold agglutinin and complement than are normal erythrocytes. The protection from complement-mediated lysis can be correlated with the membrane deposition of complement proteins. Recent studies have shed considerable light on these phenomena. It has been shown in animal models and in man that cold agglutinins mediate their effects via complement activation (26, 27). In vitro data have demonstrated that these antibodies, like other IgM antibodies studied, mediate complement activation efficiently. One molecule of cold agglutinin, bound to the erythrocyte surface, fixes one molecule of the first component of complement, C1 (28). Activation of complement by the cold-agglutinin red-cell complex proceeds in two stages. At the lower temperature, the cold agglutinin binds to the erythrocyte surfaces and binds (or "fixes") the first component of the serum complement system, C1 (30). At intermediate temperatures, C4 can bind to the erythrocyte membrane. When the temperature returns to 37° (in the body when the erythrocyte returns from acral areas, such as the fingers, toes, or tip of the nose, to the core circulation), the other components in the complement sequence interact with the erythrocyte membrane. Depending upon the number of cold-agglutinin molecules that coat the cells, the antigenic specificity of the cold agglutinin, and the conditions of exposure to complement, sufficient complement may be activated to lead to intravascular hemolysis (31). Intravascular hemolysis, however, does not usually occur. More often, insufficient complement is activated on the erythrocyte membrane to lyse the cell, although enough complement is bound to produce cell sequestration by the RES. Recent evidence indicates that IgM antibodies mediate their effect via complement activation and, more particularly, via membrane attachment of the opsonic complement fragment, which at present is thought to be C3b (26, 32). If sufficient C3b is bound to the erythrocyte surface, the cell is rapidly cleared as it traverses the sinusoids of the liver (23, 27, 33, 34). The liver is rather efficient at removing these cells, and only a small percentage survive to be removed by the spleen (32).

Earlier workers believed that complement deposition at the erythrocyte surface led to major membrane damage. It was believed that the cells were cleared from the circulation by the liver because this organ was the major site of clearance of markedly damaged cells (33, 35). This concept does not adequately explain the experimental observations.

The mechanism of clearance in fact appears to be binding of the complement-coated erythrocytes to the hepatic fixed macrophages, the Kupffer cells, by attachment to specific macrophage C3b receptors (26, 29, 32). Thus the erythrocytes with membrane-bound C3b are rapidly removed within minutes of the time that they are introduced into the circulation. These complement-coated erythrocytes, which adhere to the Kupffer cell surface, undergo one of two fates (29, 32). Some of the cells are phagocytosed. The binding of C3b to the Kupffer cell receptor is, however, an inefficient stimulus for phagocytosis (36). A clearly oversimplified but conceptually helpful framework is available for examining this reaction. There exists in serum a protein (the C3 inactivator) capable of cleaving erythrocyte-bound C3b into two fragments (37). A larger fragment, termed C3c, is released from the erythrocyte and a small fragment, C3d, remains adherent to the erythrocyte membrane. It has been shown that the C3 inactivator causes the release from macrophage C3b receptors of erythrocytes with membrane-bound C3d. Thus, in the presence of the C3 inactivator, some of the erythrocytes are released from Kupffer-cell trapping and return to the circulation. It has been demonstrated that such cells with membrane-bound C3d undergo near-normal survival (29). The presence of membrane-bound C3d explains the above-noted resistance of Coombs-positive erythrocytes from patients with cold-agglutinin syndrome to further lysis by cold agglutinin and complement. Cold-agglutinin antibody attaches to the red cell and activates C3. C3b binds to all of the local sites chemically suitable for C3b binding. These sites continue to be occupied by the cleavage fragment C3d after action of the C3b inactivator. The membrane sites suitable for further C3b binding are therefore occupied, and the erythrocyte is protected from the further lytic action of cold agglutinin. It is of interest that complement is activated and C3b binds to sites that are quite close to those of initial cold-agglutinin attachment. Although these local sites can be blocked with C3d, other sites are available on the erythrocyte surface that can interact with C3b. Thus these erythrocytes, with cold agglutinin–associated C3b sites blocked, can interact with anti-A IgM isoagglutinin; under these circumstances C3b binding now proceeds normally (25, 27). The fact that the site for binding to macrophage receptors on cold agglutinin–sensitized erythrocytes is transient and can be destroyed probably helps to explain the rather mild nature of the cold-agglutinin syndrome. If these patients can stay out of a cold environment, they usually do very well and do not require therapy. This is fortunate, since the drugs usually found effective in the treatment of autoimmune hemolytic anemia (corticosteroids and cytotoxic agents) are often generally ineffective in the treatment of cold-agglutinin syndrome (1, 3). Another recent observation in animal models may be important in understanding the pathophysiologic basis of this interesting syndrome. Although C3b is usually a poor stimulus for ingestion of C3b-coated erythrocytes bound to the C3b receptor on the macrophage surface, the situation is altered if the macrophage becomes activated. Thus an activated macrophage can bind and ingest a C3b-coated RBC (38, 39). This may help to explain the frequently noted exacerbations of this syndrome seen in these patients at the time of an intercurrent infection that can result in macrophage activation. In this setting, erythrocytes that might normally have been released from the macrophage surface, and contributed to the circulating RBC pool, are instead ingested and destroyed.

Finally, some mention should be made of current thoughts on the etiology of the cold-agglutinin syndrome. It has been well demonstrated that Ii antigens are present on cell surfaces throughout adult life (40). Theories that hypothesize the sudden availability as an immunogen of sequestered antigens are therefore unlikely. The most reasonable current hypothesis for the presence of cold agglutinins following infectious illness is that microorganisms have cross-reactive antigens that stimulate an immune response (41). It has been suggested that some strains of mice, which do not have Ii antigen as part of their makeup, synthesize anti-I following mycoplasma infection (42). If this observation is correct, there must be Ii-like material in the mycoplasma organism. A variation of this theory is that the microorganism, or a product of the microorganism, binds to a cell membrane and alters the membrane surface to make the I antigen a good immunogen. Such theories remain speculative; however, they may help to explain cold-agglutinin disease mediated by antibody of restricted heterogeneity. It has been well demonstrated in animal models that certain microorganisms may induce the synthesis of antibodies with greatly restricted heterogeneity, especially in a host with the proper genetic background (43). Studies have suggested that the anti-i antibodies formed after infectious disease have somewhat restricted heterogeneity (44). It may be that in the elderly host, such antibodies with highly restricted heterogeneity are first formed following an infection, and become dominant and form the background for the development of long-standing cold-agglutinin disease.

Literature Cited

1. Dacie, J. V. 1961. *The Haemolytic Anaemias, Congenital and Acquired: Part II. The Autoimmune Haemolytic Anemias.* New York: Grune & Stratton 2nd ed.
2. Schubothe, H. 1965. Current problems of chronic cold hemagglutinin disease. *Ann. NY Acad. Sci.* 124:484–90
3. Schubothe, H. 1966. Cold hemagglutinin disease. *Semin. Hematol.* 3:27–47
4. Harboe, M. 1971. Cold auto-agglutinins. *Vox. Sang.* 20:289–305
5. Roelcke, D. 1974. Cold agglutination: antibodies and antigens. *Clin. Immunol. Immunopathol.* 2:266–80
6. Donath, J., Landsteiner, K. 1904. Uber paroxysmale hemoglobinurie. *Muench. Med. Wochenschr.* 51:1590–1600
7. Feizi, T. 1975. Ii antigens. *Proc. R. Soc. Med.* 68:799–802
8. Finland, M., Peterson, O. L., Allen, H. E., Samper, B. A., Barnes, M. W. 1945. Cold agglutinins II: Cold isohemagglutinins in primary atypical pneumonia of unknown etiology with a note on the occurrence of hemolytic anemia in the cases. *J. Clin. Invest.* 24:458
9. Spigam, C. L., Jones, J. P. 1945. Cold hemagglutination in primary atypical pneumonia and other common infections. *Arch. Intern. Med.* 76:75–87
10. Harboe, M. 1964. Interactions between ^{131}I trace-labelled cold agglutinin, complement and red cells. *Br. J. Haematol.* 10:339–46
11. Evans, R. S., Bingham, M., Turner, E. 1965. Autoimmune hemolytic disease: Observations of serological reactions and disease activity. *Ann. NY Acad. Sci.* 124:422–40
12. Rosse, W. F., Logue, G. L., Adam, J., Crookston, J. H. 1974. Mechanism of immune lysis of the red cells in hereditary erythroblastic multinuclearity with a positive acidified serum test and paroxymal nocturnal hemoglobinuria. *J. Clin. Invest.* 53:31–43
13. Rosse, W. F., Sherwood, J. B. 1970. Cold reacting antibodies. Differences in the reaction of anti-I. Antibodies with adult and cord red blood cells. *Blood* 36:28–42
14. Fudenberg, H. H., Kunkel, H. G. 1957. Physical properties of red cell agglutinins in acquired hemolytic anemia. *J. Exp. Med.* 106:689–702
15. Roelcke, D., Ebert. W., Feizi, T. 1974. Studies on the specificities of two IgM

lambda cold agglutinins. *Immunology* 27:879–86

16. Feizi, T., Kabat, E. A. 1972. Immunochemical studies on blood groups LIV. Classification of anti-I and anti-i sera into groups based on reactivity patterns with various antigens related to the blood group a, B, H, Le^a, Le^b, and precursor substances. *J. Exp. Med.* 135:1247–58
17. Rosse, W. F., Lauf, P. K. 1970. Reaction of cold agglutinins with I antigens solubilized from human red cells. *Blood* 36:777–84
18. Williams, R. C., Kunkel, H. G., Capra, J. D. 1968. Antigenic specificities related to cold agglutinin activity of gamma M globulins. *Science* 161: 379–81
19. Lecompte, J., Feizi, T. 1975. A common idiotype on human macroglobulins with anti-I and anti-i specificity. *Clin. Exp. Immunol.* 20:287–302
20. Garratty, G., Petz, L. D., Brodsky, I., Fudenberg, H. H. 1973. An IgA high titer cold agglutinin with an unusual blood group specificity within the Pr complex. *Vox Sang.* 25:32
21. Pruzanski, W., Farid, N., Keystone, E., Armstrong, M., Greaves, M. F. 1975. The influence of homogeneous cold agglutinins on human B- and T-lymphocytes. *Clin. Immunol. Immunopathol.* 4:248–57
22. Pruzanski, W., Farid, N., Keystone, E., Armstrong, M. 1975. The influence of homogeneous cold agglutinins on polymorphonuclear and mononuclear phagocytes. *Clin. Immunol. Immunopathol.* 4:277–85
23. Evans, R. S., Turner, E., Bingham, M., Woods, R. 1968. Chronic hemolytic anemia due to cold agglutinins. II. The role of C^1 in red cell destruction. *J. Clin. Invest.* 47:691–701
24. Engelfriet, C. P., Von Den Borne, A. E., Von Den Borne, K. R., Beckers, D., Reynierse, E., Van Loghem, J. J. 1972. Autoimmune haemolytic anaemias. V. Studies on the resistance against complement haemolysis of the red cells of patients with chronic cold agglutinin disease. *Clin. Exp. Immunol.* 11: 255–64
25. Evans, R. S., Turner, E., Bingham, M. 1967. Chronic hemolytic anemia due to cold agglutinins: the mechanism of resistance of red cells to C^1 hemolysis by cold agglutinins. *J. Clin. Invest.* 46: 1461–74
26. Brown, D. L., Lachmann, P. J., Dacie, J. V. 1970. The in vivo behavior of complement-coated red cells: Studies in C6 deficient, C3 depleted and normal rabbits. *Clin. and Exp. Immunol.* 7:401–21
27. Jaffe, C. J., Atkinson, J. A., Frank, M. M. 1976. The role of complement in the clearance of cold agglutinin sensitized erythrocytes in man. *J. Clin. Invest.* 58:942–49
28. Rosse, W. F. 1968. Fixation of the first component of complement (C^{1}1a) by human antibodies. *J. Clin. Invest.* 47:2430–45
29. Atkinson, J. A., Frank, M. M. 1974. Studies on the in vivo effects of antibody: Interaction of IgM antibody and complement in the immune clearance and destruction of erythrocytes in man. *J. Clin. Invest.* 54:339–48
30. Boyer, J. T., Hinz, C. F. 1967. Complement and cold agglutinins. I. Complement coating of normal human erythrocytes by cold agglutinins. *Clin. Exp. Immunol.* 2:229–39
31. Logue, G. L., Rosse, W. F., Gockerman, J. P. 1973. Measurement of the third component of complement bound to red blood cells in patients with cold agglutinin syndrome. *J. Clin. Invest.* 52:493–501
32. Schreiber, A. D., Frank, M. M. 1972. The role of antibody and complement in the immune clearance and destruction of erythrocytes. I. In vivo effects of IgG and IgM complement fixing sites. *J. Clin. Invest.* 51:575–82
33. Lewis, S. M., Dacie, J. V., Szur, L. 1960. Mechanism of haemolysis in the cold agglutinin syndrome. *Br. J. Haematol.* 6:154–59
34. Jandl, J. H., Jones, A. R., Castle, W. B. 1957. The destruction of red cells by antibodies in man. I. Observations on the sequestration and lysis of red cells altered by immune mechanisms. *J. Clin. Invest.* 36:1428–58
35. Jandl, J. H., Kaplan, M. E. 1960. The destruction of red cells by antibodies in man. III. Quantitative factor influencing the patterns of hemolysis in vivo. *J. Clin. Invest.* 39:1145–56
36. Mantovani, B., Rabinovitch, M., Nussenzweig, V. 1972. Phagocytosis of immune complexes by macrophages. Different roles of macrophage receptor sites for complement (C3) and for immunoglobulin (IgG). *J. Exp. Med.* 135:780–92
37. Ruddy, S., Austen, K. F. 1971. C3b inactivator of man. II. Fragments pro-

duced by C3b inactivator cleavage of cell bound or fluid phase C3b. *J. Immunol.* 107:742–50

38. Atkinson, J. A., Frank, M. M. 1974. The effect of Bacillus-Calmette-Guerin. Induced macrophage activation on the in vivo clearance of sensitized erythrocytes. *J. Clin. Invest.* 53:1742–49
39. Bianco, C., Griffin, F. M. Jr., Silverstein, S. C. 1975. Studies of the macrophage complement receptor. Alteration of receptor function upon macrophage activation. *J. Exp. Med.* 141:1278–90
40. Weens, J. H., Schwartz, R. S. 1974. Etiologic factors in autoimmune hemolytic anemia. *Ser. Haematol.* 7:303–27
41. Costea, N., Yakulis, V., Heller, P. 1965. Experimental production of cold agglutinins. *Blood* 26:323–40
42. Costea, N., Yakulis, V., Heller, P. 1969. The mechanism of induction of cold agglutinins by M. pneumoniae. *Blood* 34:829 (Abstr.)
43. Krause, R. M. 1970. The search for antibodies with molecular uniformity. *Adv. Immunol.* 12:1–53
44. Jacobson, L. B., Longstreth, G. F., Edington, T. S. 1973. Clinical and immunologic features of transient cold agglutinin hemolytic anemia. *Am. J. Med.* 54:514–21

Ann. Rev. Med. 1977. 28:299–309
Copyright © 1977 by Annual Reviews Inc. All rights reserved

TREATMENT OF IDIOPATHIC THROMBOCYTOPENIC PURPURA (ITP)[1]

◆7244

Yeon S. Ahn, M.D. and William J. Harrington, M.D.
Center for Blood Diseases, Department of Internal Medicine, University of Miami School of Medicine, Miami, Florida 33152

INTRODUCTION

The term *idiopathic thrombocytopenic purpura* (ITP), as used in this review, is applied only to a disorder seen mostly in older children, adolescents, and adults. It has no known relationship to any underlying disease or any inciting factor. Once it develops, if untreated, it is chronic, recurrent or relapsing, and usually life-long in duration (even though most such patients, when first seen, give histories of acute disease, short in duration, reflecting alarm at the appearance of abnormal bleeding).

It is to be distinguished from so-called ITP of young children, a syndrome that usually follows viral infections or immunizations of many kinds and, while acute, is generally self-limited, with spontaneous recovery.

It is also to be distinguished from other thrombocytopenias that may mimic it, such as certain drug-induced thrombocytopenias, thrombocytopenia in collagen-vascular diseases, in lymphoreticular disorders, and in other malignancies.

ITP is one of the best-defined diseases of autoimmunity. Circulating autoantibodies to platelets were first demonstrated, both in vivo and in vitro, in 1951 (1, 2). They cause prompt destruction of transfused platelets (3, 4); are globulins (1, 4) of the IgG class (5); react with platelets, rendering them susceptible to destruction in the reticuloendothelial system, (especially the spleen); and can traverse the placenta, causing neonatal thrombocytopenia (6).

The antibody in ITP is seemingly directed against a platelet-specific antigen, since there is seldom concomitant autoimmune hemolytic anemia or neutropenia. When an associated hemolytic anemia occurs (7) (Evans' syndrome) (2), it probably re-

---

[1] Supported in part by Grant #AM 09001 11 and Grant #2 M01 RR0024, as well as by funds provided by the Stephen Andrew de Young Leukemia Foundation, the Joseph Peter Siegel Memorial Fund for Leukemia Research, and the Milton N. Weir, Jr., Research Fund.

flects the presence of two antibodies, one platelet-specific and one red cell–specific, inasmuch as such patients often have alternating periods of hemolytic anemia without thrombocytopenia, or thrombocytopenia without hemolytic anemia.

A few patients with so-called ITP (less than 5%) have a syndrome that in all other respects resembles autoimmune ITP (6), except that their platelet survival times are normal (6, 8), and women with this syndrome give birth to normal infants (6). This variant requires reclassification; it is presumed that it reflects a lack of megakaryocyte maturation and not an autoimmune destruction of platelets.[2]

TREATMENT

The attainment of a platelet count of 30,000/mm^3 is usually sufficient to prohibit serious bleeding in ITP, and, of 50,000/mm^3, is sufficient to assure safety. The normal platelet level, by direct counting methods, ranges between 150,000 and 350,000/mm^3.

Treatment of ITP consists of general and more specific measures. The latter include splenectomy, administration of glucocorticoids, and use of other agents that handicap macrophage function and act as immunosuppresants.

General Measures

Irrespective of the cause of thrombocytopenia, bleeding at any platelet level will be enhanced by trauma, fever, infection, increase in metabolic rate from any cause, azotemia, or the administration of agents that decrease platelet production or handicap platelet function. Also, because of an increase in intravascular pressure, bleeding tends to be more prominent in the more dependent portions of the body.

Accordingly, avoidance of trauma is important; fever should be treated with antipyretics such as acetaminophen; infections should be treated where possible. Azotemia should be corrected if possible. Causes for increase in metabolism should be rectified. The head should be kept elevated at times of most alarming bleeding. Alcohol ingestion in excess should be avoided, since it decreases platelet production (9) and viability. Most important, drugs that handicap platelet function should be avoided (10). The most notable of these is aspirin.

Although their beneficial effect on capillary fragility has been questioned (11), it is generally agreed that glucocorticoids do improve capillary fragility in all thrombocytopenic states irrespective of the cause of the thrombocytopenia (12). An average dosage in thrombocytopenic adults of 15 mg of prednisone daily is usually sufficient to bring about improvement in bleeding tendency, independent of its effect on platelet counts.

[2] Citation of the extensive literature on the pathologic physiology of ITP and on the techniques and problems of platelet serology in ITP is beyond the scope of this review on treatment. Many additional contributions have been made, most notably by J. Dausset, P. A. Miescher, M. Stefanini, W. Dameshek, F. H. Gardner, N. R. Shulman, R. H. Aster, M. Baldini, S. Karpatkin, L. A. Harker, R. I. Handin, R. Dixon, and their colleagues.)

Specific Measures

Theoretically, since ITP is a disorder of autoimmunity, the use of immunosuppressants would seem most logical. However, because of their side effects in a disease which, once established, usually requires lifelong treatment, immunosuppressants (including glucocorticoids) have their greatest value in patients who have failed to respond to splenectomy.

SPLENECTOMY Splenectomy was first used for the treatment of ITP in 1916, following the suggestion of Kaznelson that the spleen acts as a filter to remove platelets from the circulation (13). Interpreted in the light of present understanding, it appears that platelets are "sensitized" by a circulating platelet-specific IgG, making them susceptible to removal by the reticuloendothelial macrophages of the spleen. During surgery, lessening of bleeding is commonly observed the moment the splenic pedicle is ligated (12); following splenectomy, an increase in platelets is generally noted within a few hours, although it may be delayed for days or occasionally for four to five weeks. The usual promptness of remissions is such that decrease in antibody production cannot be a factor, although the spleen may additionally produce some of the autoantibody (14). However, if the platelet is severely affected by the antibody, its destruction takes place even in the absence of a spleen (1, 4) (e.g. in the liver) (15–17). It has been proposed that selective splenic sequestration of labeled platelets may be used as a means of selection of candidates for splenectomy (18, 19). This claim is not warranted (20, 21).

The overall permanent remission rate from splenectomy in ITP (as defined in this review) is in the range of 70–80% (22–30), and the platelet count six weeks after surgery is a good indicator of its future behavior; later relapse is rare (30). Certain criteria are useful in predicting responsiveness. The probability of a favorable effect is greater in patients with (*a*) less severe disease (4, 15, 31); (*b*) prior response to glucocorticoids (26, 28, 32); (*c*) age under 45 years (26); and (*d*) non-autoimmune forms of ITP (4, 6, 26, 33) (not accepted by all investigators).

The risks from splenectomy are small; in our own and other series, the mortality from surgery has been less than 1% (22–30), even though remission is not always accomplished. Accordingly, splenectomy affords a prompt and lasting clinical cure in most patients with ITP, with acceptable low risk to them. Because of its generally prompt effect, splenectomy is required as an emergency measure in patients with alarming (e.g. intracranial) bleeding.

Preparation with platelet transfusions is unnecessary and undesirable, considering the risks of hepatitis and the lack of effectiveness of platelet transfusions in nearly all patients with ITP (3, 4, 8, 34).

The presence of one or more accessory spleens was formerly considered to be a reasonably common cause for relapse after splenectomy (35, 36). However, this is very rare (37, 38). Patients dying later whose thrombocytopenia has not recurred have been found to have accessory spleens, and patients who have relapsed very rarely have accessory spleens (21). Indeed, exploratory laparotomy, with unsuccessful search for an accessory spleen, has been observed to be followed by a remission

(21). There are techniques with radionucleides that make the search for an accessory spleen feasible (38, 39). Much simpler is an examination of the peripheral blood smear for Howell-Jolly bodies and Heniz bodies, both of which are regularly present in subjects who have been totally splenectomized. Again, it is very rare that an accessory spleen is responsible for continued or recurrent thrombocytopenia.

GLUCOCORTICOIDS Glucocorticoids were first demonstrated to have effectiveness in the treatment of ITP in 1950 (12). ACTH and cortisone were then used. It was observed that a decrease in severity of purpura could commonly be achieved within one or two days, with an improvement in platelet count often within a week. Complete remissions (i.e. with normalization of platelet counts) are distinctly less common (15–60%) than those achieved with splenectomy, require several days or weeks to be attained, and relapses almost invariably occur on discontinuance of therapy (23–26).

The initial beneficial effects of glucocorticoids on platelet levels probably represent handicaps imposed upon reticuloendothelial macrophages (15), although with more protracted therapy a decrease in autoantibody synthesis also occurs (4, 40). At times, patients who had been refractory to glucocorticoid therapy prior to splenectomy and who failed also to respond to splenectomy may respond to steroids postoperatively (24).

When a patient with ITP is initially seen, glucocorticoids are generally used except in instances of critical bleeding (when splenectomy is indicated, as noted above). By use of glucocorticoids, the physician is given more time to determine more precisely the nature of the thrombocytopenia; also, though only in an occasional case, a complete remission will be obtained, which persists after the steroids are discontinued. Data cited in the literature (23–25, 28, 29) imply higher rates (14–38%) of sustained remission following discontinuance of steroids; however, these include observations on young children and adults whose thrombocytopenia is secondary to viral or other agents and is self-limited.

Also, steroids are indicated in patients in whom splenectomy is contraindicated.

For initial therapy in patients who have clinical bleeding, doses of prednisone (or its equivalent) of 1–1½ mg/kg per day are used. Therapy should be continued at this level for two to three weeks. If the platelet count then does not exceed a level of 50,000 per cubic millimeter, splenectomy should be performed. If greater improvement takes place, the dosage of glucocorticoids should be decreased to a level that can be maintained safely, [e.g. 0.25 mg/kg per day]. If relapse occurs, splenectomy should be performed. If remission is sustained, the dosage may be adjusted to a 0.25 mg/kg on alternate days or every third day.

As noted, glucocorticoid therapy is generally not satisfactory for long-range management, and therefore its use is regarded now as adjunctive in most patients.

It should also be noted that glucocorticoid therapy may actually perpetuate thrombocytopenia in some patients (41), and remission will take place only with a decrease in dosage or discontinuance of the steroid. However, such an event is very rare, and when the choice between increased or decreased dosages of steroids is to be made in the presence of serious bleeding manifestations, an increase to 300 or more milligrams of prednisone or its equivalent per day for one week should be tried first.

Moderate to large doses of androgens may induce remissions and enhance glucocorticoid responsiveness (42).

IMMUNOSUPPRESSANTS Since 1960 (43), a variety of nonsteroidal immunosuppressant agents have been used in the treatment of ITP refractory to splenectomy and glucocorticoids. In general, these agents handicap multiple aspects of the immune response, including macrophage function. With the exception of the vinca alkaloids, they require prolonged periods of treatment before an effect is usually seen (periods on the order of six weeks to three months). The most frequently employed are azathioprine, cyclophosphamide, and the vinca alkaloids.

Azathioprine (44–47) is used in dosages of 1.2–2.4 mg/per kg (50–250 mg per day), and a minimum of 3 months, and up to 26 months of therapy is required before the platelet count is stabilized. Glucocorticoids are used concomitantly. The side effects are leukopenia, alopecia, ulcerations of the mucous membranes, and diarrhea.

Cyclophosphamide is used in dosages of 50–200 mg/day (47–49) or intermittently (300–600 mg per square meter intravenously or by mouth) over a 4-hr period every 2–3 weeks) (50). Improvement in the platelet count may be evident in 2–4 weeks, with the maximum elevation in 4–6 weeks (47), but longer periods of treatment are usually needed. Side effects are severe leukopenia, alopecia, and hemorrhagic cystitis. [Although it has been advocated as an initial measure and an alternative to splenectomy (49), even in young children this approach is not recommended.] The overall response rate to azathioprine and/or cyclophosphamide in splenectomy failures (no improvement for one month after surgery) ranges from 25% (47) to 66% (49). Cyclophosphamide seems to be more effective than azathioprine, but its potential side effects are more serious (47).

The vinca alkaloids, vincristine and vinblastine (51–54), were recently introduced in the treatment of ITP. These are potent immunosuppresants of both humoral and cellular immunity (55). In contrast to most immunosuppressive agents that cause thrombocytopenia, these agents induce thrombocytosis in animals (56, 57) and man (58, 59). Vincristine is used in dosages of 0.025 mg/kg (not more than two mg intravenously in the average adult) (54); vinblastine, in dosages of 0.125 mg/kg (10 mg intravenously in the average adult). Injections of vinca alkaloids are repeated every 7–10 days (54). A rise in platelets is usually observed within 5–7 days after injection. This early response is an obvious advantage over azathioprine and cyclophosphamide. When there is no response after three injections, the likelihood of benefit is lessened and side effects are increased. Once remission occurs, no maintenance therapy is given. Therapy is usually well tolerated. Vinblastine appears to be as effective as vincristine, although the duration of remissions is perhaps shorter (42). When neuropathies become a problem in responders, vinblastine can be used instead of, or in alternation with, vincristine (42).

With vincristine, transient malaise, jaw pain, and fever may be observed shortly after injections. Various neuropathies (particularly sensory ones involving the hands and feet) are more troublesome and are seen somewhat more commonly in older people. These are reversible after discontinuation of treatment. Neutropenia is the major side effect of vinblastine therapy.

The use of vinca alkaloids should usually be reserved for patients whose thrombocytopenia is refractory to splenectomy and glucocorticoids. Among these refractory cases, partial or complete remissions were induced in 11 out of 16 instances (69%) (42, 54). They should not be used as the primary treatment unless there is a contraindication to steroids and splenectomy.

In the treatment of severely thrombocytopenic patients with vinca alkaloids, no aggravation of bleeding has been observed, even though vinca alkaloid-induced platelet dysfunctions in vitro have been reported (60).

Various other immunosuppressive agents have been tried in patients with ITP. These include chlorambucil, methotrexate, asparaginase (53), and actinomycin C (61). These have been used in only a small number of patients; their efficacy cannot be evaluated.

As noted, in some clinics immunosuppressive agents have been used in nonsplenectomized patients (48, 49). The responses are often of short duration, requiring continuation of therapy; a complete, unmaintained remission is seldom induced (47, 54). In contrast, immunosuppressive agents are much more effective when used in splenectomized patients, with a better rate of complete remission (47, 49, 54). Considering the potential risks involved with prolonged use of immunosuppressive agents (all of which are mutagenic) (62, 63), they should not be used in the primary modality of treatment. Their use in the pediatric age group and in women of childbearing age should additionally be discouraged because of the self-limited nature of childhood "ITP" and the teratogenic potential of these agents.

It has been emphasized that the outcome of immunosuppressive treatment with azathioprine and cyclophosphamide (47, 49) depends on the duration of the disease; if of short duration (less than seven months), most respond favorably, while those of long duration seldom respond. However, duration of disease also reflects the failure of other measures and thereby embodies a selecting out of more severely affected patients. Nonetheless, in the case of vinca alkaloids, favorable responses do not appear to depend upon the duration of the disease (54). A trial designed for induction of remission in patients with refractory ITP, using vinca alkaloids and maintaining with cyclophosphamide, is underway (42).

Once complete remission is obtained, every effort should be made to discontinue immunosuppressants. Long-term maintenance therapy is not recommended unless symptomatic thrombocytopenia recurs.

Neonatal Thrombocytopenia and Idiopathic Thrombocytopenic Purpura in Pregnancy

Neonatal thrombocytopenia may occur in newborns of mothers with idiopathic thrombocytopenia even if these women were "cured" months or years before by splenectomy; it may also occur de novo without the presence or history of thrombocytopenia in the mother (64). Special problems are presented by ITP in pregnancy and by neonatal thrombocytopenia. If untreated, there is a 10% maternal (65, 66), and up to a 25% infant mortality (65), whereas with proper recognition and management both are nil (6).

It is now well established that two mechanisms may underlie these forms of neonatal thrombocytopenia. Autoantibody from mothers with or without a history of idiopathic thrombocytopenia may be responsible (6, 67). Furthermore, with the discovery of platelet antigens distinct from those of the red cell (6), it was realized that isoimmunization may cause neonatal thrombocytopenia in a manner analogous to that of Rh isoimmunization in hemolytic disease of the newborn. Both platelet-specific and histocompatibility antigens may cause isoimmune neonatal thrombocytopenia (67–69).

During the first trimester of pregnancy, women with ITP should be treated with splenectomy, since glucocorticoids and nonsteroidal immunosuppressants are teratogenic during this period. The surgery may precipitate abortion (21).

Later in pregnancy, glucocorticoids may be used, but with careful attention to their side effects, including an increased susceptibility to eclampsia (70, 71). There are no adequate data on the use of other agents. Splenectomy later in pregnancy may be carried out if necessary through a transthoracic approach (successful in the two such instances tried in this clinic) (21).

However, later in pregnancy, when a viable infant can be expected, transabdominal splenectomy with cesarean section may be employed, successful in the three such instances in our experience (21) and also successful in the experience of others (72).

Most women with ITP or a history of ITP give birth to thrombocytopenic infants although they themselves have had splenectomy-induced "cures" even years before (6, 26). The infant having a spleen has thrombocytopenia in the presence of antibody. The greatest threat to the infant is birth trauma leading to intracranial bleeding. Accordingly, delivery should be accomplished with minimal instrumentation. Alternatively, if there is a history of severe neonatal thrombocytopenia from previous pregnancies, delivery should be accomplished by cesarean section.

If there was only mild neonatal thrombocytopenia in newborn infants from previous pregnancies, administration of prednisone (e.g., 1 mg/kg per day during the last two weeks of pregnancy) should be considered. There are no adequate data on the use of other immunosuppressants.

Postpartum Thrombocytopenia

Seventeen episodes of so-called ITP have been observed in eight women, occurring 4–6 weeks after delivery, miscarriage, or spontaneous abortion (21, 73, 74). In no instance has there been any relationship to medications or to other recognizable factors save for the postpartum state, and the pattern of recurrence in several women following successive pregnancies establishes the syndrome as having a relationship to the postpartum state. It therefore is not "idiopathic," strictly defined. However, since it has not been described in other series and therefore is not generally recognized as having a relationship to the postpartum state, it would probably be misclassified as ITP.

This syndrome responds to glucocorticoid therapy. Prompt remissions occur, and recurrence does not take place upon discontinuance of the steroid treatment. It does take place, however, following future pregnancies, although in the one instance in which splenectomy was carried out, subsequent pregnancies were not followied by recurrence. During pregnancy itself, the affected woman is hematologically normal, and the newborn baby is unaffected.

SUMMARY

In summary, idiopathic thrombocytopenic purpura (as defined here) should be treated initially with glucocorticoids unless there is life-threatening hemorrhage (such as intracranial bleeding), in which case emergency splenectomy is indicated, since it generally gives the most prompt improvement in platelet count of all modes of treatment.

Steroid therapy provides time for the physician to observe the course of the disease and to better define the cause of the thrombocytopenia. A complete, sustained remission following treatment with steroids may occasionally be obtained.

In most cases the steroid response is transient, and splenectomy is the most definitive measure with respect to a lifelong cure. Complete clinical cure is obtained in 70–80% of instances following splenectomy, with a very low operative risk. The remainder either fail to respond or experience relapse. In this group of patients, steroids should again be tried; if a good response is obtained, steroids should be given every second or third day to lessen long-term side effects.

If these measures fail, the use of immunosuppressive agents is justified. Both vinca alkaloids and cyclophosphamide are especially useful. Vinca alkaloids act more promptly, but cyclophosphamide may have a more lasting effect. Azathioprine may also be used; however, it appears to yield a lower remission rate. Nonsteroidal immunosuppressive agents should not be used as the primary modality of therapy. And, due to their carcinogenic and teratogenic potential, they should be avoided, if possible, especially in children and in women during their childbearing years.

Literature Cited

1. Harrington, W. J., Minnich, V., Hollingsworth, J. W., Moore, C. V. 1951. Demonstration of a thrombocytopenic factor in the blood of patients with thrombocytopenic purpura. *J. Lab. Clin. Med.* 38:1–10
2. Evans, R. S., Takahashi, K., Duane, A. B., Payne, R., Lui, C. 1951. Primary thrombocytopenic purpura and acquired hemolytic anemia: evidence for a common etiology. *Arch. Intern. Med.* 87:48–65
3. Hirsch, E. O., Gardner, F. H. 1952. The transfusion of human blood platelets. *J. Lab. Clin. Med.* 39:556–69
4. Sprague, C. C., Harrington, W. J., Lange, R. D., Shapleigh, J. B. 1952. Platelet transfusions and the pathogenesis of I.T.P. *J. Am. Med. Assoc.* 150:1193–98
5. Shulman, N. R., Marder, V. J., Weinrach, R. S. 1965. Similarities between known antiplatelet antibodies and the factor responsible for thrombocytopenia in idiopathic purpura: physiologic, serologic, and isotopic studies. *Ann. NY Acad. Sci.* 124:499–542
6. Harrington, W. J., Sprague, C. C., Minnich, V., Moore, C. V., Ahlvin, R. C., Dubach, R. 1953. Immunologic mechanisms in idiopathic and neonatal thrombocytopenic purpura. *Ann. Intern. Med.* 38:433–69

7. Fisher, J. A. 1947. Cryptogenic acquired hemolytic anemia. *Q. J. Med.* 16:245–62
8. Cohen, P., Gardner, F. H., Barnett, G. O. 1961. Reclassification of thrombocytopenias by Cr^{51}-labeling method for measuring platelet life-span. *N. Engl. J. Med.* 264:1294–99, 1350–55
9. Post, R. M., Desforges, J. F. 1968. Thrombocytopenia and alcoholism. *Ann. Intern. Med.* 68:1230–36
10. Mustard, J. F., Packham, M. A. 1975. Platelets, drugs, and thrombosis. *Drugs* 9:19–76
11. Bonnin, J. A. 1961. The management of thrombocytopenic states with particular reference to platelet thromboplastic function. I. Idiopathic and secondary thrombocytopenic purpura. *Br. J. Haematol.* 7:250–60
12. Robson, H. N., Duthie, J. J. R. 1950. Capillary resistance and adrenocortical activity. *Br. Med. J.* 2:971–77
13. Kaznelson, P. 1916. Verschwinden der hämorrhagischen Diathese bei einem Falle von "essentieller Thrombopenie" (Frank) nach Milzextirpation. *Wien. Klin. Wochenschr.* 29:1451–54
14. Karpatkin, S., Strick, N., Suskind, G. W. 1972. Detection of splenic antiplatelet antibody synthesis in idiopathic autoimmune thrombocytopenic purpura (ATP). *Br. J. Haematol.* 23: 167–76
15. Shulman, N. R., Weinrach, R. S., Libre, E. P., Andres, H. L. 1965. Role of reticulo-endothelial system in pathogenesis of idiopathic and thrombocytopenic purpura. *Trans. Assoc. Am. Physicians* 78:374–90
16. Aster, R. H., Keene, W. R. 1969. Sites of platelet destruction in idiopathic thrombocytopenic purpura. *Br. J. Haematol.* 16:61–73
17. Figallo, M. A., Harrington, W. J. 1972. Platelet production and destruction. In *Hematopoietic and Gastrointestinal Investigations with Radionuclides,* ed. A. J. Gilson, W. J. Smoak III, M. B. Weinstein, 3:21–32. Springfield, Ill: Thomas
18. Najean, Y., Ardaillou, N. 1971. The sequestration site of platelets in idiopathic thrombocytopenic purpura: its correlation with the results of splenectomy. *Br. J. Haematol.* 21:153–64
19. Ries, C. A., Price, D. C. 1974. ^{51}Cr platelet kinetics in thrombocytopenia: correlation between splenic sequestration of platelets and response to splenectomy. *Ann. Intern. Med.* 80:702–7
20. Aster, R. H. 1972. Platelet sequestration studies in man. (Annotation) *Br. J. Haematol.* 22:259–63
21. Harrington, W. J., unpublished observations
22. Watson-Williams, E. J., Macpherson, A. I. S., Davidson, S. 1958. Treatment of idiopathic thrombocytopenic purpura. *Lancet* 275:221–26
23. Carpenter, A. F., Wintrobe, M. M., Fuller, E. A., Haut, A., Cartwright, G. F. 1959. Treatment of idiopathic thrombocytopenic purpura. *J. Am. Med. Assoc.* 171:1911–16
24. Doan, C. A., Bouroncle, B. A., Wiseman, B. K. 1960. Idiopathic and secondary thrombocytopenic purpura: clinical study and evaluation of 381 cases over a period of 28 years. *Ann. Intern. Med:* 53:861–76
25. Meyers, M. C. 1961. Results of treatment in 71 patients with idiopathic thrombocytopenic purpura. *Am. J. Med. Sci.* 242:295–302
26. Harrington, W. J., Arimura, G. 1961. Immune reactions of platelets. *Henry Ford Hosp. Symp. Blood Platelets,* pp. 659–70. Boston: Little, Brown
27. Charlesworth, D., Torrance, H. B. 1968. Splenectomy in idiopathic thrombocytopenic purpura. *Br. J. Surg.* 55:437–39
28. Thompson, R. C., Moore, R. A., Hess, C. E., Wheby, M. S., Leavell, B. S. 1972. Idiopathic thrombocytopenic purpura: long-term results of treatment and the prognostic significance of response to corticosteroids. *Arch. Intern. Med.* 130:730–34
29. JiJi, R. M., Firozvi, T., Spurling, C. L. 1973. Chronic idiopathic thrombocytopenic purpura: treatment with steroids and splenectomy. *Arch. Intern. Med.* 132:380–83
30. Macpherson, A. I. S., Richmond, J. 1975. Planned splenectomy in treatment of idiopathic thrombocytopenic purpura. *Br. Med. J.* 1:64–66
31. Viala, J. J., Dechavanne, M., Ville, D. 1975. Les epreuves radioisotopiques plaquettaires permettent-elles de meilleures indications de la splenectomie au cours due purpura thrombopenique idiopathique chronique? A propos de 50 observations. *Lyon Med.* 234:419–25
32. Brennan, M. F., Rappaport, J. M., Maloney, W. C., Wilson, R. E. 1975. Correlation between response to corticosteroids and splenectomy for adult idiopathic thrombocytopenic purpura. *Am. J. Surg.* 129:490–92

33. Harrington, W. J. 1954. The clinical significance of antibodies for platelets. *Sang* 25:712–25
34. Ahn, Y. S., Harrington, W. J. 1975. Platelet transfusions in clinical medicine. *Adv. Intern. Med.* 20:379–98
35. Thorek, P., Gradman, R., Welch, J. S. 1948. Recurrent primary thrombocytopenic purpura with accessory spleen. *Ann. Surg.* 128:304–11
36. Rosenthal, N., Vogel, P., Lee, S., Lipsay, J. 1951. The role of accessory spleens in post-splenectomy recurrent purpura hemorrhagica. *J. Mt. Sinai Hosp. NY* 17:1008–20
37. Lozner, E. L. 1954. The thrombocytopenic purpuras. *Bull. NY Acad. Med.* 30:184–94
38. Loeb, V. Jr., Seaman, W. B., Moore, C. V. 1952. The use of thorium dioxide sol (Thorotrast) in the roentgenologic demonstration of accessory spleens. *Blood* 7:904–14
39. Eckelman, W., Richards, P., Atkins, H. L., Hauser, W., Klooper, J. F. 1971. Visualization of human spleen with ^{99M}Tc-labelled red blood cells. *J. Nucl. Med.* 12:310–11
40. Suhrland, L. G., Arquilla, E. R., Weisberger, A. S. 1958. The effect of prednisone on circulating antibody formation in animals immunized with human platelet antigen. *J. Lab. Clin. Med.* 51:724–30
41. Cohen, P., Gardner, F. H. 1961. Thrombocytopenic effect of sustained high-dosage prednisone therapy in thrombocytopenic purpura. *N. Engl. J. Med.* 265:611–617
42. Ahn, Y. S., Harrington, W. J. Unpublished observations
43. Dameshek, W., Schwartz, R. 1960. Treatment of certain "autoimmune" diseases with antimetabolites: a preliminary report. *Trans. Assoc. Am. Physicians* 73:113–27
44. Bouroncle, B. A., Doan, C. A. 1966. Refractory idiopathic thrombocytopenic purpura treated with Imuran. *N. Engl. J. Med.* 275:630–35
45. Sussman, L. N. 1967. Azathioprine in refractory idiopathic thrombocytopenic purpura. *J. Am. Med. Assoc.* 202: 259–63
46. Bouroncle, B. A., Doan, C. A. 1969. Treatment of refractory idiopathic thrombocytopenic purpura. *J. Am. Med. Assoc.* 207:2049–52
47. Finch, S. C., Castro, O., Cooper, M., Covey, W., Erichson, R., McPhedran, P. 1974. Immunosuppresive therapy of chronic idiopathic thrombocytopenic purpura. *Am. J. Med.* 56:4–12
48. Laros, R. K., Penner, J. A. 1971. "Refractory" thrombocytopenia purpura treated successfully with cyclophosphamide. *J. Am. Med. Assoc.* 215:445–49
49. Verlin, M., Laros, R. K., Penner, J. 1976. Treatment of refractory thrombocytopenic purpura with cyclophosphamide. *Am. J. Hematol.* 1:97–104
50. Weinerman, B., Maxwell, I., Hryniuk, W. 1974. Intermittent cyclophosphamide treatment of autoimmune thrombocytopenia. *Can. Med. Assoc. J.* 16:1100–02
51. Sultan, Y., Delobel, J., Jeanneau, C., Caen, J. P. 1971. Effect of periwinkle alkaloids in idiopathic thrombocytopenic purpura. *Lancet* 1:496–97
52. Marmont, A. M., Damasio, E. E., Gori, E. 1971. Vinblastine sulfate in idiopathic thrombocytopenic purpura. *Lancet* 2:94
53. Marmont, A. M., Damasio, E. E. 1971. Clinical experiences with cytotoxic immunosuppressive treatment of idiopathic thrombocytopenic purpura. *Acta Haematol.* 46:74–91
54. Ahn, Y. S., Harrington, W. J., Seelman, R. C., Eytel, C. 1974. Vincristine therapy of idiopathic and secondary thrombocytopenia. *N. Engl. J. Med.* 291:376–80
55. Aisenberg, A. C., Wilkes, B. 1964. Studies on the suppression of immune responses by the periwinkle alkaloids vincristine and vinblastine. *J. Clin. Invest.* 43:2394–2403
56. Hwang, Y. F., Hamilton, H. E., Sheets, R. F. 1969. Vinblastine-induced thrombocytosis. *Lancet* 2:1075–76
57. Robertson, J. H., Crozier, E. H., Woodend, B. E. 1970. The effect of vincristine on the platelet count in rats. *Br. J. Haematol.* 19:331–37
58. Robertson, J. H., McCarthy, G. M. 1969. Periwinkle alkaloids and the platelet count. *Lancet* 2:353–55
59. Carbone, P. P., Bono, V., Frei, E. III, Brindley, C. O. 1963. Clinical studies with vincristine. *Blood* 21:640–47
60. Steinherz, P. G., Miller, D. R., Hilgartner, M. W., Schmalzer, E. A. 1976. Platelet dysfunction in vincristine-treated patients. *Br. J. Haematol.* 32:439–50
61. Martin, H., Nowicki, L., Schubert, J. C. F., Schubert, H. 1967. Behandlung des morbus werlhof mit Actinomycin C. *Dtsch. Med. Wochenschr.* 92:1061–65

62. Sieber, S. M., Adamson, R. H. 1975. Toxicity of antineoplastic agents in man: chromosomal aberrations, antifertility effects, congenital malformation, and carcinogenic potential. *Adv. Cancer Res.* 22:57–155
63. Penn, I. 1974. Occurrence of cancer in immune deficiencies. *Cancer* 34:858–66
64. Epstein, R. D., Lozner, E. L., Cobbey, T. S. Jr., Davidson, C. S. 1950. Congenital thrombocytopenic purpura. *Am. J. Med.* 9:44–56
65. Robson, H. N., Davidson, I. S. 1950. Purpura in pregnancy with special reference to idiopathic thrombocytopenic purpura. *Lancet* 2:164–69
66. Heys, R. F. 1966. Child bearing and idiopathic thrombocytopenic purpura. *J. Obstet. Gynecol. Br. Commonw.* 73:205–14
67. Harrington, W. J., Minnich, V., Arimura, G. 1956. The autoimmune thrombocytopenias. *Prog. Hematol.* 1:166–92
68. Pearson, H. A., Shulman, N. R., Marder, V. J., Cone, T. E. Jr. 1964. Isoimmune neonatal thrombocytopenia purpura. *Blood* 23:154–77
69. Colombani, J., Colombani, M., Dausset, J. 1968. Two cases of neonatal thrombocytopenia due to maternal isoimmunization against leukoplatelet antigen. *Vox Sang.* 14:137–42
70. Harrington, W. J. 1957. Therapy of the purpuras. *J. Chronic Dis.* 6:365–82
71. Heys, R. F. 1966. Steroid therapy for idiopathic thrombocytopenic purpura. *Obstet. Gynecol.* 28:532–42
72. Jones, W. R., Storey, B., Norton, G., Neische, F. W. Jr. 1974. Pregnancy complicated by acute idiopathic thrombocytopenic purpura. *J. Obstet. Gynecol. Br. Commonw.* 81:330–32
73. Harrington, W. J. 1971. Differential diagnosis and management of thrombocytopenia. *Med. Times* 99:53–66
74. Harrington, W. J. 1971. Chronic idiopathic thrombocytopenic purpura. In *The Blood Platelets,* ch. 18, pp. 264–72. Baltimore: Williams & Wilkins

Ann. Rev. Med. 1977. 28:311–27
Copyright © 1977 by Annual Reviews Inc. All rights reserved

THE ROLE OF ONCOGENIC VIRUSES IN NEOPLASIA[1]

❖7245

Paul A. Klein, Ph.D. and Richard T. Smith, M.D.
Tumor Biology Unit, Department of Pathology, University of Florida
College of Medicine, Gainesville, Florida 32610

INTRODUCTION

Concerning discovery in the natural sciences, Lewis Thomas has written, "There are special kinds of information that seem to be chemotactic. As soon as a trace is released, receptors at the back of the neck are caused to tremble, there is a massive convergence of motile minds flying upwind on a gradient of surprise, crowding around the source. It is an infiltration of intellects, an inflammation" (1). Ironically, this type of response did not greet the reports of Ellerman & Bang in 1908 or of Rous in 1910 on the transmissibility of an avian leukemia and an avian sarcoma with a filterable agent (2, 3). Although Rous eventually fulfilled most of Koch's postulates, which dominated that important era of infectious disease research, his inability to induce tumors in mammals with his "tumor virus" led most of the scientific world to disregard the avian system as an anomaly unsupportive of a general viral etiology of cancer.

The virus-cancer concept began to gain support between 1930 and 1950 from the mammalian-based experiments of Shope (4), Bittner (5), Gross (6), and others (7). It was not until the 1960s, however, that the idea of a virus etiology of cancer began to attract the efforts of a critical multidisciplinary mass of virologists, geneticists, epidemiologists, immunologists, and molecular biologists. As a result, evidence has accumulated that establishes (*a*) that some naturally occurring animal cancers are actually caused by viruses, (*b*) that the ability of these viruses to cause such cancers is under genetic control of the host, and (*c*) that immunologic prophylaxis that has provided protection against many conventional virus diseases may in certain instances be effective in preventing virus-induced cancers.

As yet, no unequivocal etiologic relationship between a virus and any malignant human cancer has been established. On the other hand, a growing body of virologic, serologic, and epidemiologic evidence supports a role for certain viruses in specific human tumors.

---

[1] This is Tumor Biology Unit publication No. 112 and was supported, in part, by Grants CA-09126, CA-15334, HD-00384, CB-53933, CA-13384, from the National Institutes of Health, Bethesda, Maryland, and Grant F76UF-5 from the American Cancer Society.

Here we review the state of knowledge in this field, with special emphasis on recent virologic and immunologic data involving human cancer and animal models that appear analogous to human tumors. Several comprehensive and recent reviews are available that can supplement the virological (7–12), immunological (13–18) or multicategorical (19–21) data described briefly here.

GENERAL ETIOLOGICAL CONSIDERATIONS

Cancer in both animals and humans is acknowledged to be a multifactorial disease in which genetic background, host defenses, and environmental influences play critical roles (13, 22, 23). Thus, as a general example, the host genotype may determine the oncogenic potential of an environmentally abundant carcinogen independent of other factors. As a specific example, the enzyme aryl hydrocarbon hydroxylase (AHH) is one of the mixed-function oxidase enzyme systems that converts nonpolar polycyclic aromatic hydrocarbons to excretable water-soluble products. The chemically reactive intermediates formed by this enzyme are both cytotoxic and carcinogenic (24). In mice, the inducibility of the AHH system segregates as a single autosomal dominant gene. Strains of mice carrying this gene are easily inducible for AHH activity and highly sensitive to 3-methylcholanthrene (MCA) carcinogenesis, presumably due to the rapid and efficient metabolism of MCA to its ultimate carcinogenic form (25). Recent studies have suggested that humans who are heavy smokers and whose cells are easily inducible for AHH are more likely to develop bronchogenic carcinoma than "low" inducible types who smoke heavily (26).

Similarly, infection with a known oncogenic virus does not invariably result in tumor development. Thus the genotype of the host influences tumor-virus function at the cellular level (22, 27, 28) and affects host immunologic responsiveness directed at newly transformed tumor cells and infectious virus (13–16, 28–31, and see below).

TYPES AND FREQUENCY OF CANCER

Animal tumor systems have contributed so much of our knowledge concerning the molecular biology and pathobiology of tumor viruses that it is tempting a priori to relate animal cancer to human cancer by analogy. However, in no animal species do the incidence and variety of naturally occurring tumors duplicate those seen in man (22, 32–37, 41). Table 1, for example, compares tumors of domestic mammals with those of man in the United States. The most thoroughly studied mammal, the laboratory mouse, also develops a spectrum of spontaneous tumors different from those of man. These have been well documented for commonly used inbred strains (35, 36), as well as for several wild-mouse populations (35, 37). While several inbred laboratory strains show a very high incidence of mammary cancer (C3H) or leukemia (AKR), many of the most common human tumor types are rare (lung, digestive tract, etc) (35–37).

Table 1 Leading sites of cancer in humans and domestic animals in the United States[a]

Tumor site	Approximate incidence (%)					
	Human					
	male	female	Cattle	Horse	Cat	Dog
Breast	<1	26	<1	<1	5	13
Leukemia and lymphoma	8	7	20	2	33	8
Respiratory system	24	7	<1	<1	4	3
Bone and connective tissue	1	1	<1	<1	3	4
Genital organs	18[b]	20[c]	9	10	1	5
Buccal cavity and pharynx	5	2	1	1	8	5
Digestive organs	26	24	2	1	7	10
Urinary organs	9	4	<1	<1	2	<1
Endocrine glands	2	2	<1	<1	1	2
Brain and central nervous system	2	2	<1	4	1	2
Eye	<1	<1	49	12	2	3
Skin[d]	1	1	10	45	18	28
All other and unspecified sites	4	4	5	15	12	11

[a] Adapted from data in (32) and (34).
[b] Predominantly tumor of the prostate.
[c] Predominantly carcinoma of the uterine cervix (14%) and ovary (5%).
[d] Incidence for humans includes melanoma only; incidence for animals includes all types of skin tumors.

Viruses in the Etiology of Cancer

In this context we focus on the evidence and the possibilities that at least some animal and human cancer is actually caused by the action of oncogenic viruses. Most of the naturally occurring virus-induced tumors of animals are associated with the RNA viruses. These cause mammary carcinoma, leukemias, and sarcomas in the mouse, and leukemias and sarcomas in the cat and chicken (20 and see below). The DNA tumor viruses, including herpesviruses, papovaviruses, and adenoviruses, also induce sarcomas, lymphomas, and papillomas, but usually in hosts of species different from the natural host of the virus. *H. saimiri,* for example, is a naturally occurring non-oncogenic virus of the squirrel monkey, but induces a rapidly lethal reticulum-cell sarcoma in owl monkeys.

The notion that every virus-induced animal tumor must have a virus-induced human counterpart, however attractive, is unsupported. Even if each type of proven virus-induced cancer in animals actually had an anatomic analog and virus-induced counterpart in man, the total could account only for about one fourth of all human cancers (Table 1), and would not include such medically significant forms as cancer of the lung, colon-rectum, esophagus-stomach, prostate, or liver-pancreas. Clearly, appropriate workable animal models are lacking for some of the most important forms of human cancer (38–40).

In the absence of animal models, then, what types of evidence can be developed for or against a contention that a virus is actually the etiologic agent of a human cancer? First, epidemiologic and serologic studies may either define a clear correlation between the occurrence of infection by a suspected tumor virus and a particular tumor, or demonstrate case clustering in time or geography, which is suggestive of an infectious process. Second, electron microscopic, biochemical, virological, and immunological techniques that detect tumor viruses in animal systems may be used to seek evidence of analogous viral activity in human tumors and normal human tissues. Third, immunoprophylaxis (vaccination) with vaccines prepared from suspected tumor viruses might achieve a reduction in the incidence of particular tumors and thus support an etiologic role for those viruses in types of specific human cancer. Fourth, examination of those patients having deficiency disorders involving antiviral defenses might reveal the occurrence of unusual tumors.[2] Each of these approaches has been utilized or is under way.

CHARACTERISTICS OF THE KNOWN TUMOR-ASSOCIATED VIRUSES

The general division of tumor-associated viruses into DNA and RNA classes is based upon the character of the genetic material of the complete virus particle. However, both general classes of viruses have the capacity to intrude into the host genome directly or indirectly as DNA sequences. The properties of the known RNA virus are discussed first.

RNA TUMOR VIRUSES

The oncornaviruses (a term derived from oncogenic RNA viruses) (42) have proven oncogenic potential expressed in the naturally occurring leukemias, sarcomas, and mammary carcinomas of several species (43–45). The intricate genetic "modus operandi" of the oncornaviruses has both inspired and perplexed the search for human oncogenic viruses during the past decade. Oncornaviruses are distributed widely among the vertebrates and share common structural, biochemical, physical, and immunological features (12, 42–45). These viruses are spherical, enveloped particles, 80–110 nm in diameter. An electron-dense nucleoid contains the viral genome within a ribonucleoprotein complex, while the surface envelope possesses the short antigenic glycoprotein spikes or knobs responsible for adsorption of the virus to a host cell. Two major subgroups are distinguished morphologically by the location of their nucleoid within the virus core: the type C particles with a centrally positioned nucleoid and the B particles with an eccentrically located nucleoid. Both type C and B particles are released from cells by budding through the cell membrane. A third morphologic subgroup, type A, has smaller particles with a double outer shell, which are seen either in the cytoplasm or around the nucleus. The relationship of type A particles to types B and C is unclear. Type C particles are usually found associated with animal leukemias and sarcomas, whereas the type B particles are exclusively associated with mammary carcinoma.

---

[2] The emphasis in this pursuit up to now has been upon supporting a surveillance role for the immune system against tumor cells, rather than selective susceptibility to oncogenic viruses of human or animal origin (13, 113).

Oncornaviruses are comprised of six major structural proteins that are designated by the prefix *p* (protein) or *gp* (for glycoproteins), followed by the abbreviated molecular weight of the protein (45). For example, p30 refers to the major 30,000-MW internal protein, which comprises 10–12% of the total protein of the virus, while gp70 refers to the 70,000-MW glycoprotein spike found on the envelope of the murine oncornaviruses. The molecular weights of analogous oncornavirus proteins vary little from one virus to another, but each viral protein has unique immunologic specificities. These specificities provide the means for "tracking" individual viruses or virus proteins during animal studies and in searching for oncornaviruses of authentic human origin (43, 44, 45). Antigenic specificities currently known include type-specific antigens, which are carried by a particular virus strain; group-specific or species-specific antigens, which are common to all oncornaviruses isolated from a given host species; and interspecific determinants, which are shared by viruses from two or more different mammalian species (43, 45). In addition to structural proteins, the viral genome codes for an RNA-dependent DNA polymerase, or reverse transcriptase (p60), which can synthesize homologous DNA from a single-stranded RNA template (46, 47). This ability to convert viral RNA to DNA permits integration of the viral genome with that of the host cell (46, 47).

The genome of the oncornaviruses is polyploid, i.e. consists of 3–4 identical RNA subunits, each with an approximate molecular weight of 3×10^6 (44, 48). The complete oncornavirus genome is also referred to as a *virogene* (see below) (49–51). The total genetic information in each subunit can encode about 300,000 daltons of amino-acid sequence, most of which has already been identified in the form of viral structural proteins (45, 48). Four major genetic "loci" have been mapped within the genome of the avian sarcoma viruses (48): The *onc* locus or *oncogene* (50, 51) regulates the transforming or oncogenic potential of the virus; the *pol* locus specifies the RNA-dependent DNA polymerase (p60 or "reverse transcriptase"); the *env* locus specifies the major envelope glycoprotein of the virus (gp 70); and the *gag* locus encodes other major structural proteins of the virus (p10, 12, 15, 30) (45, 48).

When a virogene is integrated into susceptible host cell DNA and the viral genome contains an oncogene (*onc*), the growth properties of the infected cell are altered and the infected cells become transformed, behaving in vitro and in vivo like malignant cells. The onc locus is thought to specify the synthesis of a yet-unidentified transforming protein with a molecular weight of up to 50,000 (48).

The transforming potential of leukemogenic and sarcomogenic oncornaviruses is usually studied using viruses obtained from tumors (Rous sarcoma virus; Gross, Friend, Rauscher, or Moloney leukemia viruses, etc) (7, 43, 44). These viruses behave like ordinary viral infectious agents when injected into either newborn or adult animals, replicating to high titers and causing tumors (7). Recently it has become apparent that the oncogenic capability of the widely used laboratory strains is not a regular attribute of the oncornaviruses. Thus, type C particles found in many normal tissues and even in tumors are usually not oncogenic (53). This has led to the disclosure that virogenes are present in one or more copies in each somatic and germ cell of each individual among many species (49–51). These virogenes are transmitted from parent to progeny according to Mendelian expectations (49–53). Activation of virogene expression within otherwise normal cells or tumor cells may occur either spontaneously, or following diverse cell stimuli. This results in synthesis and release of type C infectious oncornaviruses in the absence of malignant transformation. The *onc* locus or oncogene function is therefore not required for replication of oncornaviruses (48–51). Agents that can induce expression of endogenous oncornaviruses include analogs of some of the most widely used cancer chemotherapeutic agents (82). Graft-versus-host reactions are also known to activate oncornavirus production (95).

It is postulated that oncornaviruses may have evolved from animal cells in which they perform some unknown but vital functions during normal differentiation or adult life (43, 49, 54–56). Furthermore, they may have served as natural transmitters of genes between the species during evolution (49). Most important to our considerations, they do have the capability to transmit oncogenic information (i.e. oncogenes) both horizontally (infectious form) and vertically (parent to offspring through the gametes). In either case, they can disrupt normal cellular controls and cause neoplasia in several species.

Nontransforming (i.e. non-oncogenic) endogenous oncornaviruses released from cells do not usually cause cytopathic changes in the cells they infect. They are detected by electron microscopy, by measuring the presence of their viral DNA polymerase [reverse transcriptase], by measuring antigenic viral proteins in cells and tissues, by nucleic-acid hybridization using labeled viral RNA or labeled DNA transcripts of the viral RNA, and in some cases by tissue-culture virus-plaque assays (43). On these various grounds, the murine endogenous oncornaviruses are divided into three tentative subgroups. The first includes those that replicate only in cells of the species of origin (ecotropic); the second includes those that are restricted or unable to grow in the species of origin, but can grow in cells of many other species (xenotropic) (57). The third includes those that replicate well in both the species of origin and other species (amphotropic) (58). The ecotropic murine leukemia viruses are subdivided further into N and B tropic groups on the basis of whether they can infect cells from certain genetically defined strains of mice (59).

Certain avian sarcoma viruses, termed "helper-independent," induce tumors that produce infectious oncogenic virus. Sarcomas induced by these viruses grow both by tumor cell division and the continuous transformation of additional normal cells by the infectious virus. In contrast, all known mammalian type C sarcoma viruses (i.e. feline, mouse, woolly monkey) are oncogenic in their own species, but are "defective," in that infectious progeny viruses are not produced by newly transformed sarcoma cells. Defective sarcoma virus genomes in mammals can be rescued by superinfection of cells with nontransforming type C oncornaviruses (i.e. leukemia viruses). The leukemia "helper" virus provides the defective sarcoma virus with the necessary envelope proteins it lacks for formation of complete infectious virus. The progeny resulting from such rescue of defective sarcoma viruses are of two types: a pseudotype virus having the sarcoma genome contained within an envelope or coat provided by an alien rescuing type C leukemia virus. The second virus produced is the infectious leukemia helper virus. An important consequence of this helper effect is that the pseudotype virus assumes the host range of the helper virus that rescued it. Thus, for example, if a defective murine sarcoma virus is rescued by ecotropic mouse leukemia helper virus, the resulting progeny replicate only in mouse cells. However, if the same murine sarcoma virus is rescued by a feline leukemia virus helper (FeLV), the pseudotype sarcoma virus grows in cat, dog, and human cell cultures—the normal host range for FeLV (43), but not in mouse cells. This phenomenon has been widely used as a test for the presence of putative type C helper viruses (nontransforming) and has been an important tool for recovering defective transforming oncornaviruses from tumor cells.

Oncornaviruses and Leukemia

Leukemia is transmitted horizontally as a classical infectious disease by oncornaviruses in chickens and domestic cats (7, 60). Among inbred laboratory mice, murine leukemia viruses are acquired chiefly by vertical transmission, but horizontal transmission has been difficult to rule out (7). Naturally occurring leukemia in the mouse involves at least three independent loci that influence the occurrence of the disease (27). The *V* loci are chromosomal sites of leukemia virus integration; the *Fv-1* locus controls the replication of the virus, and the *Ir* region of the MHC complex controls immunological responsiveness to the virus and to some tumor-associated antigens (13, 25, 29–31). Recent trappings of wild mice have uncovered populations with a high (20%) incidence of leukemia combined with a high level of spontaneous expression of type C virus, but it is still unclear whether the virus is endogenous or is spread by horizontal infection (37).

No compelling epidemiologic evidence currently exists that either a horizontally or vertically transmitted virus is involved in any human leukemia. However, scattered clinical reports continue to sustain a modest level of enthusiasm for the notion that an oncogenic virus may be involved in some way. For example, Thomas et al report in vivo leukemic transformation of normal human marrow cells following their engrafting into an irradiated human recipient (61). Similarly, a monozygous twin with leukemia had leukocytes that contained oncornavirus-like nucleic acid not detectable in cells taken from the healthy child (62).

Most workers investigating a possible viral etiology of human leukemia are searching for infectious type-C particles or components of oncornaviruses in human leukemic and normal cells using biochemical, immunological, and virological techniques developed with mouse, cat, and primate oncornaviruses (63–71). The results obtained with human leukemias have been intriguing, confusing, and contradictory. Antigens related to the reverse transcriptase (72) and to the p30 structural protein of a woolly monkey virus (65) were detected in acute human myelogenous leukemic cells. Subsequently, a type C virus was reported to be isolated from cultured human acute myelogenous leukemic cells (64, 68, 69). The latter virus, termed HL23V, turned out to be a mixture of two distinct viruses, one indistinguishable from the woolly monkey virus (SSAV-1), the other identical to an endogenous virus of the baboon species, *Papio cynocephalus* (BALV-M7). The possibility was considered that these two viruses resulted from accidental laboratory contamination, but the fact that antigens of these viruses could be detected in fresh cells from the patient seemed to rule out this possibility (110). Since no DNA sequences related to the woolly monkey viruses have been found in normal human cells (7, 100), it is assumed that these viruses were horizontally transmitted to man (70, 73, 110, 111).

Several groups have independently reported the isolation of type C particles related to the woolly monkey virus from human tissues and tumors (66, 67). These data argue that primate viruses may be more widespread in human population than had been suspected. However, the argument for horizontal infection is not supported by other studies (70), in which no evidence could be found for the presence of woolly monkey viruses in over 200 individual tissue cultures of human origin. Nor was antibody reactivity found to be directed against woolly monkey p30 or gp70 in over 400 normal human sera and sera from patients with tumors. These apparently contradictory results cannot be resolved with the available data.

Additional evidence supporting an association between the oncornaviruses and some human leukemias has been obtained by Metzgar and co-workers (71). Highly specific antisera prepared against purified structural proteins of the Friend murine leukemia virus (FLV) reacted with some human leukemia cells (CML, AMML), but not with others (AML, ALL, CLL). Normal peripheral leukocytes did not react. Their data indicate that whatever the mechanism of leukemogenesis is in man, CML and AML cells possess membrane antigens that cross-react with FLV structural components, including p30 and gp70 (71).

Oncornaviruses and Mammary Carcinoma

Oncornaviruses with a type B morphology are the etiologic agents of mammary adenocarcinoma in inbred mice (5, 7, 74). The mammary tumor virus (MTV) is transmitted in milk, seminal fluid, the gametes, and can be detected in its integrated form in most tissues of all mouse strains using specific nucleic-acid hybridization techniques (75). However, the incidence of mammary cancer in different mice varies widely among different strains, being strongly influenced by genetic factors. In contrast to other oncornaviruses, satisfactory in vitro assays of MTV biological activities are not available.

Type B particles resembling MTV of mice are only rarely found in human milk and the presence of these particles does not correlate with a propensity to develop breast cancer (74). Other epidemiologic data is not consistent with a horizontally transmitted infectious agent in human breast cancer. Some evidence supporting the existence of a human analog of mouse MTV is found in nucleic-acid hybridization experiments. These show the presence of mouse MTV nucleic-acid–like sequences in human breast carcinoma but not in normal breast tissue (76). Further evidence for the similarity of a putative human MTV and the mouse virus is provided by the fact that sera from 50% of breast cancer patients, 40% of relatives of some of those patients, but only 15% of normal donors reacted positively in immunofluorescence tests with mouse mammary tumor cells (77). One candidate MTV virus (734B) isolated from a tissue-culture line of human mammary adenocarcinoma has been recently characterized (78). Recent hybridization experiments suggest that this virus is of possible endogenous human origin since it anneals only with human, but not with nonhuman (including primate), nuclear DNA.

Oncornaviruses and Sarcomas

Oncornaviruses can induce sarcomas in mice, birds, and cats (43). In humans, sarcomas are infrequent and have not been studied intensively in this context. Relatives and contacts of patients with osteogenic sarcoma have a significantly higher incidence of demonstrable antibody and cell-mediated immunity to osteogenic sarcoma antigens than the normal population (79–81). In one study, 80–90% of relatives of osteogenic sarcoma patients had antisarcoma antibodies as measured by immunofluorescence, in contrast to 29% of the normal population (77, 78). These data suggest, but do not prove, involvement of a transmissible infectious agent in such tumors.

Oncornaviruses and Human Cancer in Perspective

The epidemiologic studies of human cancer fail to support any general etiologic role for viruses acting as horizontally transmitted infectious agents (7). To oncornavirus advocates, this indicates that the multiple genetic restrictions on viral oncogenesis, known in animals, are operational in man as well. It is also argued that the oncogenes of vertically transmitted endogenous human oncornaviruses are the critical participants in all types of naturally occurring or carcinogen-induced human cancer (50, 51). Indeed, mammary carcinoma or leukemia has occurred in several members of a family, over several successive generations (7), and the existence of so-called cancer families could be interpreted as evidence for a vertical mode of transmission.

Chemical carcinogens play an important and in some cases proven etiologic role in certain human cancers (19, 112). These include cancer of the lung, bladder, stomach, and liver (19). Chemical carcinogens appear not to transform via the activation of oncornavirus genomes (82). Critical interpretation of these experiments, however, is limited by our inability to measure or regularly recover integrated viral *oncogenes* from normal or transformed cells that fail to display other evidence of endogenous oncornavirus activity (82).

Other experiments, however, support a possible role for oncornaviruses in chemical carcinogenesis. These include successful immunization against chemical carcinogenesis with a vaccine prepared from species-specific type C particles (83), reduction in the incidence of chemically induced tumors with long-term interferon therapy (84), and prevention of viral-chemical cocarcinogenesis with neutralizing antibodies to oncornaviruses (85).

A proven etiologic role for oncornaviruses in human cancer would present great challenges in immunoprophylaxis and immunotherapy. In animal systems, abundant evidence has accumulated that oncornavirus-associated or -induced antigens can be immunogenic in the species of origin and that these antigens can induce under certain conditions potent tumor-rejection reactions (17, 28, 86–89). Vaccination against horizontally transmitted or induced leukemia with viral vaccines has already produced encouraging results (87, 90). Similarly, the incidence of naturally occurring vertically transmitted leukemia and mammary carcinoma in mice can be reduced in certain instances with vaccines prepared from viral antigens (17). It is known that patients with metastatic cancer and acute leukemia have immune responsiveness to oncornaviruses (91). Chemotherapy in man might properly be aimed at increasing the expression of oncornavirus antigens on tumor cell membranes, thereby enhancing the immunogenicity of otherwise weakly immunogenic tumors (49, 88).

In conclusion, a role for oncornaviruses in human cancer is currently unproven. However, they remain active candidates in cancer etiology and pathogenesis and deserve attention also because of their possible association with other human diseases. These include certain neurological disorders (92) and autoimmune diseases (93, 94). Finally, since oncornaviruses are known to be activated by allograft reactions in histocompatible graft recipients, they may be involved in certain immunopathologic processes as well as in the increased incidence of lymphoid neoplasia observed in immunosuppressed graft recipients (95).

DNA TUMOR VIRUSES

Three groups of DNA-containing viruses have a proven oncogenic potential in animal systems, the herpesviruses, adenoviruses, and papovaviruses (derived from *pa*pilloma, *po*lyoma, *va*cuolating virus) (7–10, 12, 20, 96). The papovaviruses are the etiologic agents for human warts and are associated with other human papillomatous tumors (7). The adenoviruses have not yet been associated with any type of human cancer. The herpesviruses have been implicated in the etiology of Burkitt's lymphoma, nasopharyngeal carcinoma, and cervical carcinoma (97). The three virus groups possess a diversity of structural, virological, and immunological characteristics, but are united by a common ability to integrate their DNA directly with that of their host cells, and to transform normal cells into tumor cells. Detailed discussion of the molecular biology of these three virus agents is available elsewhere (8, 10, 97–99).

Papovaviruses

Virions in this group are nonenveloped, 45 nm in diameter, and contain a DNA genome of molecular weight 3×10^6 (8, 10, 98, 99). The principal members are the polyoma virus, simian virus 40 (SV40), and the papilloma viruses. Polyoma and SV40 produce infectious viruses in cells that support their complete replication cycle (permissive cells), culminating in the release of thousands of virus particles and cell death. Alternatively, they may transform so-called nonpermissive cells; here the virus does not multiply, but does cause expression of genetic information that results in characteristic tumor-associated changes (99).

The small size of their viral genomes and their ability to transform cells with high efficiency have made papovaviruses the most thoroughly studied of the DNA tumor viruses. Molecular characterization of their genomes has progressed to a point where the complete nucleic acid base sequence of both polyoma and SV40 will soon be available (10).

Polyoma virus is a naturally occurring, widespread tumor virus of mice (7). It is so named because of its ability to induce a broad histogenetic range of tumors. It causes tumors in mice only when injected into newborns and is not known to be involved in any human cancer.

The discovery of SV40 occurred during the production of the poliomyelitis vaccine in Rhesus monkey kidney cells. It was realized at that time that many batches of both the killed and the live attenuated vaccines contained monkey kidney cell-derived infectious SV40 particles. Subsequent studies revealed that SV40 could induce tumors when injected into newborn hamsters and that it could transform such cells in vitro (8). Studies of the cancer incidence in persons inoculated with SV40-contaminated vaccine, however, did not show an increase in cancer type or frequency.

Papilloma viruses cause warts and other papillomatous tumors in humans (7, 100). Warts are troublesome to the patient but are benign tumors in terms of neoplastic potential. Laryngeal papillomas of children present more serious medical problems, but on the basis of localization rather than malignant character. On the other hand, papovaviruses have been isolated from central nervous system tissue of patients with progressive multifocal leukoencephalopathy (PML) and shown to induce brain tumors in newborn hamsters (101). This prevalent human tumor virus deserves much more extensive investigation.

Adenoviruses

Adenoviruses are much larger and more complex than the papovaviruses. Each non-enveloped particle contains a genome with a molecular weight of 25 X 10^6. However, small fragments of adenovirus DNA can transform cells in vitro and this transforming segment is not greater than 10^6 daltons (102). Adenovirus infections are very common in human populations and most humans contain antibodies to one or more of the 31 known serotypes (8). Approximately 13 of the known human serotypes have oncogenic potential in whole-animal or in vitro animal-cell systems (8, 98). Types 12, 18, and 31 are most oncogenic for the newborn hamster (8). Adenoviruses have no proven role in human tumors.

Herpesviruses

Herpesviruses are the largest and structurally most complex of the oncogenic DNA viruses. The particles are enveloped, 180 nm in diameter and contain a genome of molecular weight 10^8 (97). Purified virions contain at least 33 species of polypeptides ranging in weight from 25,000 to 280,000 (97). Herpesviruses have a proven oncogenic potential in several different vertebrate species, including the domestic fowl (Marek's disease lymphoma), the leopard frog (Lucké's renal adenocarcinoma), as well as marmosets and owl monkeys (reticulum-cell sarcomas induced by *H. saimiri* and *H. ateles*) (10, 103, 104). The target cell for transformation in both Marek's disease and *H. saimiri* infections is the T lymphocyte and the disease caused by each can be prevented by immunization with viral vaccines (103–105).

At least five different herpesviruses infect humans: herpes simplex 1 (HSV-1), herpes simplex 2 (HSV-2), herpes zoster virus (HZV), Epstein-Barr virus (EBV), and cytomegalovirus (CMV). HSV-2 is epidemiologically associated with cervical carcinoma, while EBV is associated with Burkitt's lymphoma and nasopharyngeal carcinoma (97). EBV is the etiologic agent for infectious mononucleosis (106).

The association of HSV-2 with cervical carcinoma is based first on seroepidemiological findings of a significantly higher incidence of neutralizing antibody to HSV-2 in patients with cervical carcinoma than in control populations (97, 107). Second, human cervical tumor cells contain HSV-2 DNA, viral RNA, and viral antigens (107). Third, HSV-2 can transform cells in vitro (108, 109).

Likewise, strong epidemiological, virological, and immunological evidence supports but does not prove the role of EBV as the etiological agent of African Burkitt's lymphoma (ABL) and nasopharyngeal carcinoma (NPC) (109). Almost all biopsies contain the EBV genome, whereas non–African Burkitt's lymphomas do not but can be infected by it (109).

EBV has a very restricted host range and in contrast to HSV-1 and HSV-2, causes a productive lytic infection only under restricted conditions (97). Of the two populations of human lymphocytes, EBV infects only the B- or immunoglobulin-producing cells. Furthermore, whereas normal human leukocytes cannot be maintained for long in culture, infection of such cells with EBV transforms them into permanent lines of cells. EBV, which can "immortalize" human cells, can be recovered from Burkitt's lymphoma cells and from the throat washings of patients with infectious mononucleosis. The question of whether the EB virus DNA is actually integrated into host-cell DNA and if so, in how many copies in transformed cells, is unresolved (97).

DNA Tumor Viruses in Perspective

The most significant established medical attribute of the DNA tumor viruses currently implicated in the etiology of malignant human cancer (EBV and HSV-2) is their infectious transmission from individual to individual with a benign outcome in the overwhelming majority of such infections. If the tumors presumed to be induced by these viruses indeed represent rare outcomes of such primary infections, then host elements (immunologic resistance and genetic background) as well as cocarcinogenic factors must play critical roles in their development. In the case of EBV and ABL, for example, holoendemic malaria has been proposed, but is not established as an important environmental determinant (97, 109). Prevention of these types of cancer could be directed either toward prophylaxis against primary virus infection, or elimination of cofactors in susceptible populations.

Numerous studies have clearly established virus vaccines as safe and effective in preventing a variety of herpesvirus-induced animal tumors (114). Development of analogous immunological control measures for HSV-2 in human cervical carcinoma will require the development of appropriate animal models for use in preliminary vaccine trials (40). Furthermore, a clearer understanding of the role of antibody and cell-mediated immunity in persistent and recurrent human HSV-1 and HSV-2 infections is needed (114). Several available animal models, however, could be used to study immunological prophylaxis of EBV infections for later application in man. Thus, primary EBV infection in cotton-topped marmosets results in clinically inapparent immunologic response, a lingering, self-limited, lymphoproliferative disorder, or B-cell lymphomas. In this model, problems of virus subtypes, cross-reactions, etc, could be objectively evaluated. Similarly, immunization with inactivated *H.*

saimari prevents the rapidly fatal lymphoma regularly induced in marmosets by this agent (105, 115). Applications of such animal studies to man will have the minimal requirements that any immunizing agents developed be noninfectious [i.e. DNA-free] (115), and that the immunologic balance that contains EBV-transformed cells in vivo is not upset, thus favoring escape of malignant clones (116). Prevention of ABL would eliminate the major cause of cancer in children of tropical zones. Prevention of NPC would eliminate one of the most common malignant tumors of individuals of southern Chinese origin. Lastly, effective immunization against EBV could eliminate infectious mononucleosis (106, 117), which, next to streptococcal disease, has the major responsibility for morbidity among infectious diseases of young people in the United States. Major programs along these lines should be expected in the next five years.

Literature Cited

1. Thomas, L. 1974. The lives of a cell: Notes of a biology watcher. In *Natural Science*, p. 117. Toronto, New York & London: Bantam. Paperback ed. 181 pp.
2. Ellerman, V., Bang, O. 1908. Experimental Leukemie bei Huhnern. *Centralbl. Bakteriol. Parasitenkd. Infectionskr.* 46:595–609
3. Rous, P. 1910. A transmissible avian neoplasm (sarcoma of the common fowl). *J. Exp. Med.* 12:696–705
4. Shope, R. E. 1933. Infectious papillomatosis of rabbits. *J. Exp. Med.* 58:607–24
5. Bittner, J. J. 1936. Some possible effects of nursing on the mammary gland tumor incidence in mice. *Science* 84:162–63
6. Gross, L. 1951. "Spontaneous" leukemia developing in C3H mice following inoculation in infancy with Ak-leukemia extracts or Ak-embryos. *Proc. Soc. Exp. Biol. Med.* 76:27–32
7. Gross, L. 1970. *Oncogenic viruses.* London & New York: Pergamon. 991 pp.
8. Tooze, J., ed. 1973. *The Molecular Biology of Tumor Viruses.* New York: Cold Spring Harbor Lab. 699 pp.
9. Tooze, J., Sambrook, J., eds. 1974. *Selected Papers in Tumor Virology.* Cold Spring Harbor Lab. 1026 pp.
10. Cold Spring Harbor Symposia on Quantitative Biology. 1974. *Tumor Viruses,* 39:1–1200. New York: Cold Spring Harbor Laboratory
11. Rapp, F., Westmoreland, D. 1975. Do viruses cause cancer in man? *Ca-A Cancer J. Clin.* (Am. Cancer Soc.) 25:215–29
12. Prevost, J. M. 1976. Oncogenic viruses: Implications for human diseases. Pt. I. *Eur. J. Cancer* 12:327–40
13. Smith, R. T., Landy, M., eds. 1975. *Immunobiology of the Tumor-Host Relationship.* New York & London: Academic. 360 pp.
14. Schultz, J., Leif, R. C., eds. 1975. *Critical Factors in Cancer Immunology.* New York & London: Academic. 368 pp.
15. Crowell, R. L., Friedman, H., Prier, J. E. 1976. *Tumor Virus Infections and Immunity.* Baltimore: Univ. Park Press. 271 pp.
16. Notkins, A., ed. 1975. *Viral Immunology and Immunopathology.* New York & London: Academic. 271 pp.
17. Levine, P. H., Pearson, G. R., Herberman, R. B., Rabson, A. S., Burton, G. J., eds. 1976. Immunological control of virus-associated tumors in man: Prospects and problems. *Cancer Res.* 36:559–869
18. Mathe', G., Weiner, R., eds. 1974. Investigation and stimulation of immunity in cancer patients. *Recent Results Cancer Res.* 47:1–501
19. Becker, F. F., ed. 1975. *Cancer 1: A Comprehensive Treatise, Etiology: Chemical and Physical Carcinogenesis.* New York & London: Plenum. 524 pp.
20. Becker, F. F., ed. 1975. *Cancer 2: A Comprehensive Treatise, Etiology: Viral Carcinogenesis.* New York & London: Plenum. 439 pp.
21. Gottlieb, A. A., Plescia, O. J., Bishop, D. H. L., ed. 1975. *Fundamental Aspects of Neoplasia.* New York, Heidelberg & Berlin: Springer. 424 pp.
22. Heston, W. E. 1975. Genetics: Animal tumors. See Ref. 19, pp. 33–57

23. Knudson, A. G. Jr. 1975. Genetic influences in human tumors. See Ref. 19, pp. 59–74
24. Grover, P. L., Sims, P., Huberman, E., Marquardt, H., Kuroki, T., Heidelberger, C. 1972. In vitro transformation of rodent cells by K-region derivatives of polycyclic hydrocarbons. *Proc. Natl. Acad. Sci.* 68:1098–1101
25. Kouri, R. E., Ratrie, H. III, Whitmire, C. 1974. Genetic control of susceptibility to 3-methylcholanthrene-induced subcutaneous sarcomas. *Int. J. Cancer* 13:714–20
26. Kellerman, G., Shaw, C. R., Luyten-Kellerman, M. 1973. Aryl hydrocarbon hydroxylase inducibility and bronchogenic carcinoma. *N. Engl. J. Med.* 289:934–37
27. Rowe, W. P. 1973. Genetic factors in the natural history of murine leukemia virus infection: G. H. A. Clowes Memorial lecture. *Cancer Res.* 33:3061–68
28. Chieco-Bianchi, L., Colombatti, A., Collavo, D., Sendo, F., Aoki, T., Fischinger, P. J. 1974. Tumor induction by murine sarcoma virus in AKR and C58 mice: Reduction of tumor regression associated with appearance of Gross leukemia virus pseudotypes. *J. Exp. Med.* 140:1162–79
29. Lilly, F. 1971. The influence of H-2 type on Gross virus leukemogenesis in mice. *Transplant. Proc.* 3:1239–41
30. Petranyi, G. G., Kiessling, R., Povey, S., Klein, G., Herzenberg, L., Wigzell, H. 1976. The genetic control of nature killer cell activity and its association with in vivo resistance against a Moloney lymphoma isograft. *Immunogenetics* 3:15–28
31. Nowinski, R. C. 1976. Genetic control of natural immunity to ecotropic mouse leukemia viruses: Immune response genes. *Infect. Immun.* 13:1098–1102
32. Cancer Statistics. 1976. *Ca-A Cancer J. Clin.* (Am. Cancer Soc.) 26:14–29
33. Dawe, C. J., Harshbarger, J. C., eds. 1969. A symposium on neoplasms and related disorders of invertebrate and lower vertebrate animals. *Natl. Cancer Inst. Monogr. 31,* Natl. Cancer Inst. 772 pp.
34. Priester, W. A., Mantel, N. 1971. Occurrence of tumors in domestic animals. Data from 12 United States and Canadian Colleges of Veterinary Medicine. *J. Natl. Cancer Inst.* 47:1333–44
35. Murphy, E. D. 1966. Characteristic tumors. In *Biology of the Laboratory Mouse.* New York & London: McGraw-Hill. 706 pp.
36. Smith, G. S., Walford, R. L., Mickey, M. R. 1973. Lifespan and incidence of cancer and other diseases in selected long-lived inbred mice and their F1 hybrids. *J. Natl. Cancer Inst.* 50:1195–1213
37. Gardner, M. B., Henderson, B. E., Estes, I. D., Rongey, R. W., Casagrande, J., Pike, M., Huebner, R. I. 1976. The epidemiology and virology of C-type virus-associated hematological cancers and related diseases in wild mice. *Cancer Res.* 36:574–81
38. Clark, W. H. Jr. 1976. Epidemiological implications of animal leukemia virus models for the human situation: Formal discussion. See Ref. 17, pp. 589–90
39. Brennan, M. J. 1976. Murine and rat mammary tumors as models for the immunological study of human breast cancer. See Ref. 17, pp. 728–33
40. Munoz, N. 1976. Model systems for cervical cancer. See Ref. 17, pp. 792–93
41. Campbell, J. G. 1969. Tumors of the fowl. Philadelphia: Lippincott. 292 pp.
42. Nowinski, R. C., Old, L. J., Sarkar, N. H., Moore, D. H. 1970. Common properties of the oncogenic RNA viruses (oncornaviruses). *Virology* 42:1152–57
43. Lieber, M. M., Todaro, G. J. 1975. Mammalian type C RNA viruses. See Ref. 20, pp. 91–130
44. Bishop, J. M., Varmus, H. E. 1975. The molecular biology of RNA tumor viruses. See Ref. 20, pp. 3–48
45. Strand, M., Bilello, J. A., Shapiro, S. Z., August, J. T. 1976. Genetic expression of mammalian RNA tumor viruses. See Ref. 15, pp. 1–34
46. Temin, H. M. 1976. The DNA provirus hypothesis. *Science* 192:1075–80
47. Baltimore, D. 1976. Viruses, polymerases, and cancer. *Science* 192:632–36
48. Baltimore, D. 1975. See Ref. 10, pp. 1187–1200
49. Todaro, G. J. 1976. Type C virogenes: Genetic transfer and interspecies transfer. See Ref. 15, pp. 35–44
50. Huebner, R., Todaro, G. I. 1969. Oncogenes of RNA tumor viruses as determinants of cancer. *Proc. Natl. Acad. Sci.* 64:1087–94
51. Todaro, G. J., Huebner, R. 1972. The viral oncogene hypothesis: New evidence. *Proc. Natl. Acad. Sci.* 69: 1009–15
52. Todaro, G. J., Benveniste, R. E., Callahan, R., Lieber, M. M., Sherr, C. J. 1975. Endogenous primate and feline

type C viruses. See Ref. 10, pp. 1159–68

53. Chattopadhyay, S. K., Lowy, D. R., Teich, N., Levine, A. S., Rowe, W. P. 1975. Qualitative and quantitative studies of AKR-type murine leukemia viruses sequences in mouse DNA. See Ref. 10, pp. 1085–1101
54. Temin, H. M. 1974. Relationships among RNA tumor viruses and between RNA tumor viruses and cells: Implication for the origin of viruses, the evolution of vertebrates and cellular genetic systems. *Ann. Rev. Genet.* 8:155–78
55. Jaenisch, R. 1976. Germ line integration and Mendelian transmission of the exogenous Moloney leukemia virus. *Proc. Natl. Acad. Sci.* 73:1260–64
56. Gillespie, D., Gallo, R. C. 1974. RNA processing and the origin and evolution of RNA tumor viruses. *Science* 188:802–11
57. Levy, J. A. 1973. Xenotropic viruses: Murine leukemia viruses associated with NIH Swiss, NZB, and other mouse strains. *Science* 182:1151–53
58. Hartley, J. W., Rowe, W. P. 1976. Naturally occurring murine leukemia viruses in wild mice: Characterization of a new "amphotropic" class. *J. Virol.* 19:19–25.
59. Hartley, J. W., Rowe, W. P., Huebner, R. J. 1970. Host range restrictions of murine leukemia viruses in mouse embryo cell cultures. *J. Virol.* 5:221–25
60. Essex, M. 1975. Horizontally and vertically transmitted oncornaviruses of cats. *Adv. Cancer Res.* 21:175–248
61. Thomas, E. D., Bryant, J. I., Buckner, C. D., Cleft, R. A., Fefer, A., Johnson, F. L., Nieman, P., Ramberg, R. E., Storb, R. 1972. Leukaemic transformation of engrafted human marrow cells *in vivo*. *Lancet* 1:1310–13
62. Spiegelman, S., Axel, R., Baxt, W., Kufe, D., Schlom, J. 1974. Human cancer and viral animal oncology. *Cancer* 34:1406–20
63. Gallo, R. C., Gallagher, R. E., Miller, N. R., Mondal, H., Saxinger, W. C., Mayer, R. J., Smith, R. G., Gillespie, D. H. 1974. Relationships between components in primate RNA tumor viruses and in the cytoplasm of human leukemic cells: Implications to leukemogenesis. See Ref. 9, pp. 933–61
64. Gallagher, R. E., Gallo, R. G. 1975. Type C RNA tumor virus isolated from cultured human acute myelogenous leukemia cells. *Science* 197:350–54
65. Sherr, C. J., Todaro, G. J. 1975. Primate type C virus p30 antigen in cells from humans with acute leukemia. *Science* 1187:855–57
66. Panem, S., Prochownik, E. V., Reale, F. R., Kirsten, W. H. 1975. Isolation of type C virions from a normal human fibroblast strain. *Science* 189:297–98
67. Nooter, K., Aarssen, A. M., Bentvelzen, P., DeGroot, F. G., Van Pelt, F. G. 1975. Isolation of infectious C-type oncornavirus from human leukemic bone marrow cells. *Nature* 256: 595–96
68. Okabe, H., Gilden, R. V., Hatanaka, M., Stephenson, J. R., Gallagher, R. E., Gallo, R. C., Tronick, S. R., Aaronson, S. A. 1976. Immunological and biochemical characterization of type C viruses isolated from cultured human AML cells. *Nature* 260:264–66
69. Chan, E., Peters, W. P., Sweet, R. W., Ohno, T., Kufe, D. W., Spiegelman, S., Gallo, R. C., Gallagher, R. E. 1976. Characterization of a virus (HL23V) isolated from cultured acute myelogenous leukemic cells. *Nature* 260:266–68
70. Stephenson, J. R., Aaronson, S. A. 1976. Search for antigens and antibodies crossreactive with type C viruses of the woolly monkey and gibbon ape in animal models and in humans. *Proc. Natl. Acad. Sci.* 73:1725–29
71. Metzgar, R. S., Mohanakumar, T., Bolognesi, D. P. 1976. Relationships between membrane antigens of human leukemic cells and oncogenic RNA virus structural components. *J. Exp. Med.* 143:47–63
72. Gallagher, R. E., Todaro, G. J., Smith, R. G., Livingston, D. M., Gallo, R. C. 1974. Relationship between reverse transcriptase from human acute leukemic blood cells and primate type-C viruses. *Proc. Natl. Acad. Sci.* 71: 1309–13
73. Benveniste, R. E., Heinemann, R., Wilson, G. L., Callahan, R., Todaro, G. 1974. Detection of baboon type-C viral sequences in various primate tissues by molecular hybridization. *J. Virol.* 14: 56–67
74. Moore, D. H. 1975. See Ref. 20, pp. 131–67; Ref. 21, pp. 55–59
75. Varmus, H. E., Quintrell, N., Medeiros, N., Bishop, J. M., Nowinski, R. C., Sarkar, N. H. 1973. Transcription of mouse mammary tumor virus genes in tissues from high and low tumor incidence mouse strains. *J. Mol. Biol.* 79:663–79

76. Spiegelman, S., Axel, R., Schlom, J. 1972. Virus-related RNA in human and mouse mammary tumors. *J. Natl. Cancer Inst.* 48:1205–11
77. Priori, E. S., Anderson, D. E., Williams, W. C., Dmochowski, L. 1972. Immunological studies on human breast carcinoma and mouse mammary tumors. *J. Natl. Cancer Inst.* 48: 1131–35
78. McGrath, C. M., Furmanski, P., Russo, J., McCormick, J. J., Rich, M. A. 1976. 734B: A candidate human breast cancer virus. See Ref. 14, pp. 63–87
79. Morton, D. L., Malmgren, R. A. 1968. Human osteosarcomas immunologic evidence suggesting an associated infectious agent. *Science* 162:1279–81
80. Eilber, F. R., Morton, D. L. 1970. Immunologic studies of human sarcomas: Additional evidence suggesting an associated sarcoma virus. *Cancer* 26:588–96
81. Gainor, B. J., Forbes, J. T., Enneking, W. F., Smith, R. T. 1976. Specific antigen stimulated lymphocyte proliferation in osteosarcoma. *Cancer* 37: 743–50
82. Rapp, U. R., Nowinski, R. C., Reznikoff, C. A., Heidelberger, C. 1975. Endogenous oncornaviruses in chemically induced transformation. 1. Transformation independent of virus production. *Virology* 65:392–409
83. Whitmire, C. E., Huebner, R. J. 1972. Inhibition of chemical carcinogenesis by viral vaccines. *Science* 177:60–61
84. Salerno, R. A., Whitmire, C. E., Garcia, I. M., Huebner, R. J. 1972. Chemical carcinogenesis in mice inhibited by interferon. *Nature* 239:31–32
85. Price, P. J., Bellew, T. M., King, M. P., Freeman, A. E., Gilden, R. V., Huebner, R. J. 1976. Prevention of viral-chemical co-carcinogenesis *in vitro* by type-specific anti-viral antibody. *Proc. Natl. Acad. Sci.* 73:152–55
86. Sato, H., Boyse, E. A., Aoki, T., Iritani, C., Old, L. J. 1973. Leukemia-associated transplantation antigens related to murine leukemia virus. *J. Exp. Med.* 138:593–606
87. Hunsmann, G., Moennig, V., Shafer, W. 1975. Properties of mouse leukemia viruses. IX. Active and passive immunization of mice against Friend leukemia with isolated viral GP71 glycoprotein and its corresponding antiserum. *Virology* 66:327–29
88. Greenberger, J. S., Aaronson, S. A. 1973. In vivo inoculation of RNA C-type viruses inducing regression of experimental solid tumors. *J. Natl. Cancer Inst.* 51:1935–38
89. Shellam, G. R., Knight, R. A., Mitchison, N. A., Gorczynski, R. M., Maoz, A. 1973. The specificity of effector T cells activated by tumors induced by murine oncornaviruses. *Transplant. Rev.* 29:249–76
90. Yohn, D. S., Olsen, R. G., Schaller, J. P., Hoover, E. A., Mathes, L. E., Heding, L., Davis, G. W. 1976. Experimental oncornavirus vaccines in the cat. See Ref. 17, pp. 646–51
91. Hersh, E. M., Gutterman, J. U., Mavligit, G., Gschwind, C. R., Freireich, E. J., Levine, P. H., Plata, E. I. 1974. Human immune response to active immunization with Rauscher leukemia virus. I. Cell-mediated and cell-associated immunity. *J. Natl. Cancer Inst.* 53:317–25
92. Gardner, M. B., Henderson, B. E., Officer, J. E., Rongey, R. W., Parker, J. C., Oliver, C., Estes, J. D., Huebner, R. J. 1973. A spontaneous lower motor neuron disease apparently caused by indigenous type-C RNA virus in wild mice. *J. Natl. Cancer Inst.* 51:1243–54
93. Mellors, R. C., Mellors, J. W. 1976. Antigen related to mammalian type-C RNA viral p30 proteins is located in renal glomeruli in human systemic lupus erythematosus. *Proc. Natl. Acad. Sci.* 73:233–37
94. Levy, J. 1974. Autoimmunity and neoplasia: The possible role of C-type viruses. *Am. J. Clin. Pathol.* 62:258–78
95. Melief, C. J. M., Datta, S., Louie, S., Schwartz, R. S. 1974. Immunologic activation of murine leukemia viruses. *Cancer* 34:1481–87
96. Dulbecco, R. 1976. From the molecular biology of oncogenic DNA viruses to cancer. *Science* 192:437–40
97. Roizman, B., Kieff, E. D. 1975. Herpes simplex and Epstein-Barr viruses in human cells and tissues: A study in contrasts. See Ref. 20, pp. 241–322
98. Rapp, F., Jerkofsky, M. A. 1975. DNA viruses: Molecular biology. See Ref. 20, pp. 209–39
99. Khoury, G., Salzman, N. P. 1975. Replication and transformation by papovaviruses. See Ref. 20, pp. 343–427
100. Blank, H. 1959. Common viral diseases of the skin.. *Med. Clin. North Am.* 43:1401–13
101. Walker, D. L., Padgett, B. L., Zurhein, G. M., Albert, A. E. 1973. Human papovavirus (Jc): Induction of brain tumors in hamsters. *Science* 181:674–76

102. Graham, F. L., Abrahams, P. J., Mulder, C., Heijnerker, H. L., Warnaar, S. O., de Vreis, F. A. J., Fiers, W., Van der Eb, A. J. 1974. Studies on *in vitro* transformation by DNA and DNA fragments of human adenoviruses and simian virus 40. See Ref. 10, pp. 637–50
103. Powell, P. C., Payne, L. N., Frazier, J. A., Rennie, M. 1974. T-lymphoblastoid cell lines from Marek's disease lymphomas. *Nature* 251:79–80
104. Ablashi, D. V., Pearson, G. R. 1974. *Herpesvirus saimiri,* a non human primate model for herpesvirus associated neoplasia in man. *Cancer Res.* 34: 1232–36
105. Laufs, R., Steincke, H. 1975. Vaccination of non-human primates against malignant lymphoma. *Nature* 253:71–72
106. Smith, R. T., Bausher, J. C. 1972. Epstein-Barr virus infection in relation to infectious mononucleosis and Burkitt's lymphoma. *Ann. Rev. Med.* 23:39–56
107. Aurelian, L., Strand, B. C., Jacobs, R. P., Bell, R. B., Smith, M. F. 1976. Herpesvirus antigens and cell mediated immunity in cervical cancer. See Ref. 15, pp. 89–131
108. Garfinkle, B., McAuslan, B. R. 1974. Transformation of cultured mammalian cells by viable herpes simplex virus subtype 1 and 2. *Proc. Natl. Acad. Sci.* 71:220–24
109. Klein, G. 1974. Studies on the Epstein-Barr virus genome and the EBV-determined nuclear antigen in human malignant disease. See Ref. 10, pp. 783–90
110. Reitz, M. S., Miller, N. R., Wong-Staal, F., Gallagher, R. E., Gallo, R. C., Gillespie, D. H. 1976. Primate type-C virus nucleic acid sequences (woolly monkey and baboon viruses) in tissues from a patient with acute myelogenous leukemia and in viruses isolated from cultured cells of the same patient. *Proc. Natl. Acad. Sci.* 73:2113–17
111. Wong-Staal, F., Gillespie, D., Gallo, R. C. 1976. Proviral sequences of baboon endogenous type C RNA virus in DNA of human leukemic tissues. *Nature* 262:190–94
112. Saffiotti, U., Wagoner, J. K., eds. 1976. Occupational carcinogenesis. *Ann. NY Acad. Sci.* 271:1–457
113. Smith, R. T., Landy, M., eds. 1970. *Immune Surveillance.* New York: Academic. 536 pp.
114. Parks, W. P., Rapp, F. 1975. Prospects for herpesvirus vaccination—Safety and efficacy considerations. *Prog. Med. Virol.* 21:188–206
115. Epstein, M. A. 1976. Epstein-Barr virus —is it time to develop a vaccine program? *J. Natl. Cancer Inst.* 56:697–700
116. Bausher, J. C., Smith, R. T. 1973. Studies of the Epstein-Barr virus-host relationship: Autochthonous and allogeneic lymphocyte stimulation by lymphoblast cell lines in mixed cell culture. *Clin. Immunol. Immunopathol.* 1:270–81
117. Miller, G. 1975. Epstein-Barr herpesvirus and infectious mononucleosis. *Prog. Med. Virol.* 20:84–112

intratubular pressure (6) are decreased after 24 hr of obstruction, along with RBF, probably due to afferent arteriolar vasoconstriction (5). This glomerular-vascular response to obstruction occurs at the level of the individual nephron, since reduction in glomerular filtration pressure was also noted in single tubules blocked with oil 24 hr previously, but not in adjacent unblocked tubules (7).

Partial unilateral ureteral occlusion for several hours, with acute hydronephrosis, causes a reduction in urinary sodium concentration, sodium excretion, and free-water clearance per 100 ml GFR, as well as higher urine osmolality from the obstructed kidney during water diuresis (8). The similarity of these changes to those of renal-artery constriction suggested decreased renal perfusion affecting distal nephron function. Decreased sodium excretion during partial ureteral obstruction is caused in part by increased tubular reabsorption of sodium, rather than simply by decreased filtered load (9). Increased renal tubular permeability to creatinine, mannitol, sucrose, and iothalamate, but not inulin, has been observed when proximal intratubular pressure exceeds 30 mm Hg (normal 12 mm Hg) for a brief period (10).

There is no postobstructive diuresis after relief of UUO, that is, no absolute increase in salt and water excretion (4, 9, 11). Urine flow and sodium excretion are usually considerably decreased compared to the contralateral normal kidney, although fractional excretion is markedly increased in the presence of a GFR and renal plasma flow (RPF), which are 18% and 33% of normal, respectively, in the rat (11) and 42% and 55% of normal in the dog (2). Jaenike (5) found that glomerular filtration pressure and proximal intratubular pressure were both markedly reduced after relief of 24-hr UUO and that vasodilators such as papaverine and acetylcholine, or volume expansion with mannitol or saline resulted in improved glomerular filtration. Preglomerular vasoconstriction thus appeared to be responsible for the impaired filtration. Fractional reabsorption of sodium and water, measured by micropuncture, is enhanced in the proximal and distal tubules of surface nephrons in the unilateral postobstructive kidney (11, 12) and, along with the decrease in individual-nephron GFR, results in decreased delivery of sodium to the distal nephron. Since fractional sodium excretion in the urine is increased, the impaired reabsorption must occur in the collecting duct or late distal tubule, or possibly in the juxtamedullary nephrons.

Nephron functional heterogeneity may also be important in determining the function of the postobstructive kidney. Using the ferrocyanide-microdissection (modified Hanssen's) technique, Harris & Yarger (11) found that filtration was present in 40% of superficial and in only 12% of juxtamedullary nephrons, while, using microspheres, the outer cortical blood flow was decreased and deep cortical flow relatively increased (2). It was suggested that these changes in intrarenal distribution of filtration and perfusion could be explained by preglomerular (afferent arteriolar) vasoconstriction in outer cortical nephrons, possibly secondary to renin release, and by postglomerular (efferent arteriolar) vasodilatation in the juxtamedullary nephrons, perhaps secondary to decreased blood viscosity resulting from the marked decrease in medullary osmolality of the postobstructive kidney (2).

ACUTE BILATERAL OBSTRUCTION

Intrarenal pressure remains high (30 mm Hg) during 24-hr bilateral ureteral obstruction (BUO), rather than falling to below normal values (9.2 mm Hg) as in UUO (6). This difference in degree and duration of elevation of intrarenal pressure between BUO and UUO may be important in determining the extent of subsequent nephron loss. Total RBF has not been determined prior to relief of 24-hr BUO, but there is decreased fractional perfusion of the outer cortex and a relative increase in perfusion of the deep cortex (13).

The occurrence of a massive natriuresis and diuresis after relief of complete or severe partial urinary-tract obstruction, referred to as postobstructive diuresis (POD), is a clinical event of considerable interest and importance that was first noted many years ago (14, 15). Relief of 24-hr BUO in the rat produces a similar phenomenon, in which the intrinsic renal defect has recently been studied in detail (6, 16–18). As in the clinical cases, the striking diuresis and natriuresis after BUO in the rat is not influenced by administration of vasopressin or aldosterone (6). Glomerular filtration rate is 10–20% of normal, total RBF is normal (16) or relatively higher than GFR (6), and intrarenal distribution of blood flow is normal (13, 16). Decreased glomerular filtration cannot be attributed to persisting tubular obstruction, since proximal intratubular pressure is normal (16). GFR does not increase with renal vasodilatation, and the finding of decreased renal extraction with increased hippurate clearance relative to inulin clearance suggests the presence of vasodilatation. Thus Jaenike (16) concluded that the decreased nephron GFR after relief of BUO was largely the result of efferent glomerular arteriolar vasodilatation with consequent reduction in glomerular-capillary hydrostatic pressure. Several studies indicate that surface-nephron GFR is decreased less than total-kidney GFR, suggesting more marked impairment of filtration in deep juxtamedullary nephrons (6, 16, 17). However, the presence of considerable heterogeneity of nephron function (16) makes this conclusion hazardous, and studies using the ferrocyanide-microdissection technique do not indicate consistent redistribution of nephron filtration rates during POD (19).

Tubular reabsorption of sodium and water is markedly decreased in the kidney undergoing POD. In the proximal tubule of surface nephrons, micropuncture measurements have shown normal (5, 16) or reduced (17) reabsorption. Reduced reabsorption was not associated with decreased peritubular capillary protein concentration (17). In the distal tubule, sodium and water reabsorption are consistently reduced (6, 17), but McDougal & Wright also noted that the fraction of filtered sodium remaining at the end of surface distal tubules was less than the fraction of filtered sodium in the final urine, suggesting more marked impairment of sodium reabsorption in the deep nephrons or entry of sodium into the collecting duct (17). In addition, these authors demonstrated abnormal permeability of the proximal and distal tubules to mannitol and inulin, which could be a contributory factor in the reduced fluid reabsorption along the length of the nephron during POD by allowing increased back-diffusion of sodium and water into the tubule lumen.

Medullary collecting-duct function was recently studied directly by microcatheterization in the kidney undergoing POD, and Sonnenberg & Wilson (18) observed absence of reabsorption and actual net entry of sodium and water into the collecting duct.

POSTOBSTRUCTIVE DIURESIS (POD)

Clinical studies during POD after relief of complete or severe partial obstruction of both kidneys or of a solitary kidney, have suggested that extracellular-fluid volume expansion, urea osmotic diuresis, or other circulating natriuretic factors, are possible extrarenal mechanisms accounting for POD, while impaired reabsorption in the proximal and/or distal tubule have been suggested as the important intrinsic factors in the decreased tubular reabsorption of sodium and water. Recent animal studies using intravenous reinfusion of urine (20) or cross-circulation (21) provide strong evidence that circulating natriuretic factors, particularly urea, are of major importance in the POD seen after relief of BUO but not UUO (Figure 1). Natriuretic factors other than urea, which have been demonstrated in serum and urine in chronic uremia (22–24), may also play a role in this phenomenon. Using normal rats with 24-hr complete urine reinfusion to simulate BUO, it was found that urea and circulating natriuretic factors have their major effect on inhibition of tubular reabsorption in nephron segments proximal to the medullary collecting duct (18). Accordingly, there is decreased fractional reabsorption of sodium and water in the distal and possibly proximal tubules after relief of BUO (6, 16, 17), while after relief of UUO in the absence of uremia, fractional reabsorption in these nephron segments is actually enhanced (11, 12). Because of the marked decrease in GFR after relief of BUO, the absolute amount of salt and water delivered to the medullary collecting duct was reduced in rats undergoing POD, although the urinary excretion of salt and water was markedly increased. The medullary collecting duct, therefore, appeared to be the critical nephron segment affected in the postobstructive kidney, and direct measurements of sodium and water reabsorption indicated net entry of salt and water (18). In another group of rats with 24-hr UUO there was no POD, but with urine reinfusion from the contralateral normal kidney the diuresis and natriuresis after relief of UUO was similar to that after BUO. Moreover, net addition of sodium and no significant water reabsorption were demonstrated in the medullary collecting duct of such animals. Sonnenberg & Wilson (18) concluded that the net entry of sodium into the medullary collecting duct during POD was probably a direct effect of obstruction, since it was found during POD after relief of BUO or UUO, but not with urine reinfusion alone (Figure 2).

The major mechanisms determining POD after relief of complete obstruction therefore appear to be (*a*) inhibition of proximal and distal tubular fluid reabsorption by urea and possibly other circulating natriuretic factors that accumulate in uremia; and (*b*) the direct effect of urinary-tract obstruction, perhaps through increased intrapelvic pressure, causing a marked decrease in reabsorption or actual net entry of sodium and water into the collecting ducts.

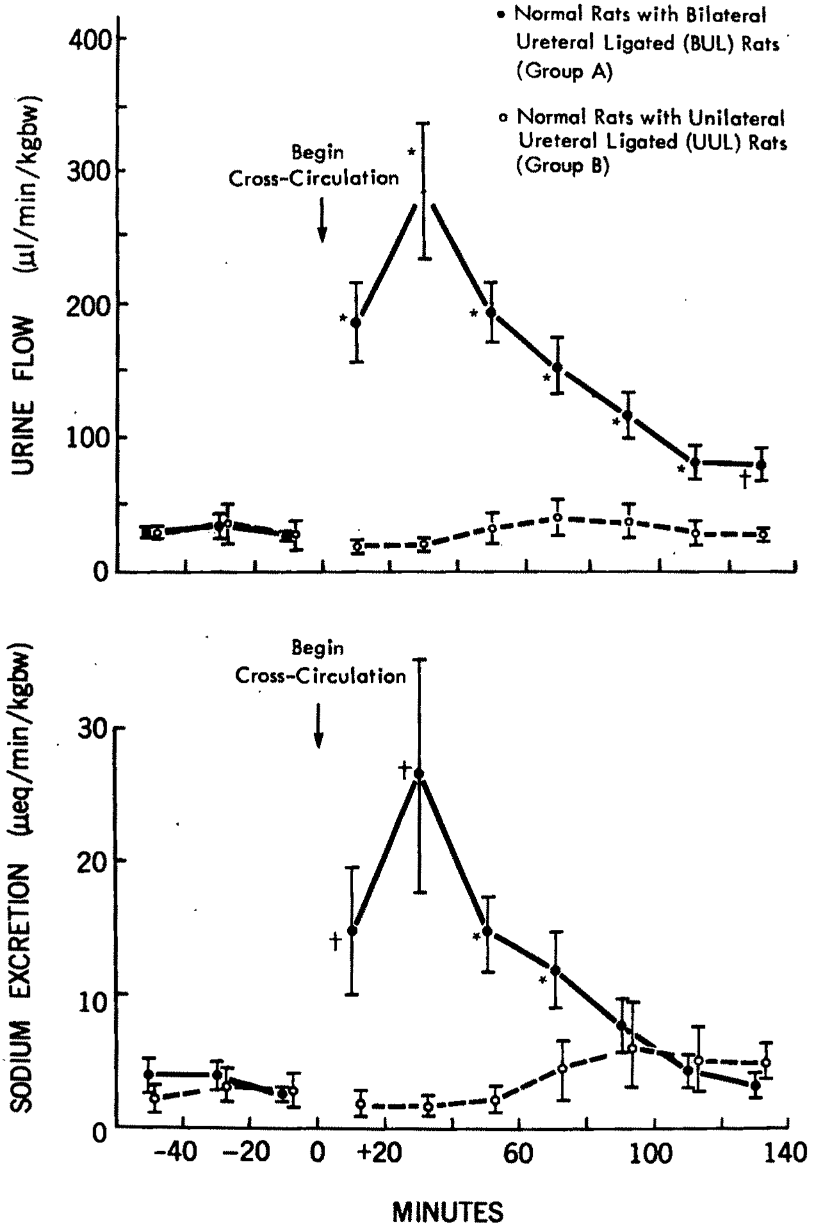


Figure 1 Changes in urine flow (*upper panel*) and sodium excretion (*lower panel*) in normal rats undergoing cross-circulation with donor rats having bilateral (●) or unilateral (○) ureteral ligation of 24-hr duration. Standard error of mean value is shown, and significance of the difference from the mean control value. * : <0.01 or less; ‡ : $P < 0.05$. (Reprinted from *J. Clin. Invest.* 57:383, with permission.)

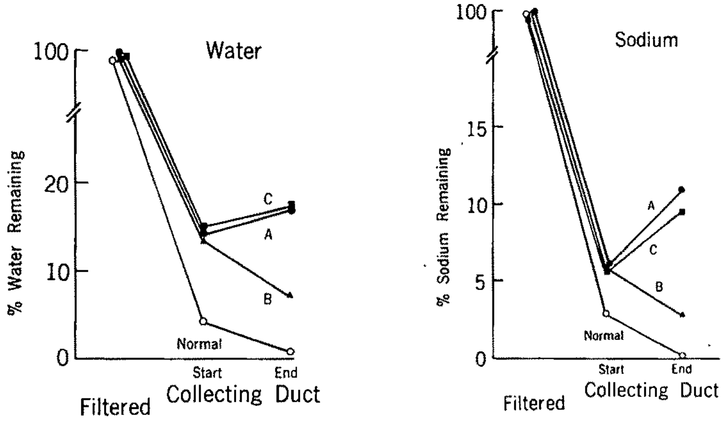


Figure 2 Percentage of filtered water and sodium remaining at beginning and end of medullary collecting duct in groups A—after relief of 24-hr bilateral ureteral ligation; B—after 24-hr total urine reinfusion intravenously (no obstruction); and C—after 24-hr unilateral ureteral ligation with urine reinfusion from the contralateral normal kidney, compared with normal sham-operated rats (Reprinted from *J. Clin. Invest.* 57:1572, with permission).

CHRONIC OBSTRUCTION

During chronic unilateral obstruction, either complete or partial, both RBF and GFR are reduced, as already noted after 24 hours of obstruction, and there is a compensatory increase in these parameters in the contralateral normal kidney. Clearance studies in chronically hydronephrotic dogs during hypotonic saline loading indicated that, in contrast to acute hydronephrosis, fractional free-water clearance and sodium excretion were increased in the hydronephrotic kidney (8). Micropuncture studies in a rat model of chronic partial obstruction have demonstrated that surface-nephron GFR was reduced, but was less impaired than whole-kidney GFR, and that proximal tubular reabsorption was normal or enhanced (25). The results suggest that impaired reabsorption in the distal tubule, deeper nephrons, or the collecting ducts was largely responsible for the decreased concentrating ability and enhanced fractional excretion of sodium. Disproportionate reduction of GFR in the juxtamedullary nephrons of chronically obstructed kidneys has also been observed more directly (19).

After relief of chronic partial obstruction, increased fractional excretion of sodium and water is seen in man (26) and in the rat model (25), but there is no postobstructive diuresis. Thus, as with acute unilateral obstruction, factors other than the intrinsic renal damage must be essential for the development of this phenomenon. Hypotonic volume expansion resulted in a disproportionate diuresis from the hydronephrotic kidney, which was associated with reduced reabsorption in the distal nephron (27). Similar increased excretion from the postobstructive kidney has been observed with mannitol administration after relief of 24-hr unilateral ureteral ligation in the dog (28). Since extracellular-fluid volume expansion is a common clinical feature of patients with complete or severe partial obstruction (14, 15, 29, 30), these results suggest that volume expansion could be a significant factor in POD, even in the absence of uremia, by exaggerating a defect in water reabsorption in the distal nephron and/or in the deep nephrons or collecting ducts of the hydronephrotic kidney.

The role of uremia and reduced nephron mass in determining the function of the chronically obstructed kidney has been studied in micropuncture experiments carried out before and after relief of chronic partial obstruction in a solitary kidney (31). Modest POD occurred with little change in surface-nephron function, indicating again the importance of impaired reabsorption in the deeper nephrons or collecting ducts. A similar increase in salt-water excretion was not seen in remnant kidneys with a similar GFR but with an intact renal medulla. High blood-urea levels were associated with more marked POD and with decreased water reabsorption in the distal nephron.

RECOVERY OF RENAL FUNCTION AFTER OBSTRUCTION

Recovery of renal function after relief of obstruction is an important clinical problem. In dogs after 7 days of UUO, there is immediate recovery of filtration to 25% of normal, with maximum improvement of 65% of control values (32). After 14 days of complete UUO, GFR and tubular function returned to 46% of the control value by 4 months after relief of obstruction and did not change thereafter (33). Sodium excretion and free-water reabsorption were similar to controls when corrected for GFR. In animals there is little evidence of improvement in function if obstruction has been present for more than four weeks.

Few human studies are available that document the degree of recovery of renal function after well-defined periods of obstruction. Recovery of function, based on roentgenographic studies, is probable for man with less than three weeks ureteral obstruction (34) and no return of function occurs after 7.5 months (35). A recent careful report indicates that recovery of renal function to 20% of normal for one kidney, a GFR of 10 ml/min, occurred after relief of 3 months of complete unilateral ureteral obstruction (36). Using these three points in time (3 weeks, 3 months, and 7.5 months), an approximation of the rate of nephron loss with persisting obstruction can be obtained. The completeness of the obstruction and subsequent rise in intrarenal pressure is probably a critical factor, as well as the presence or absence of infection. More study of suitable human cases is obviously required.

URINE ACIDIFICATION AND PHOSPHATE EXCRETION

Impaired urine acidification, as indicated by inability to lower urine pH normally in response to acidemia, has been observed in patients after relief of unilateral (36, 37), bilateral (38), or solitary-kidney (30) obstruction. In the presence of massive natriuresis, proximal bicarbonate-wasting renal tubular acidosis may be seen (30), but most studies suggest a form of distal renal tubular acidosis. After relief of UUO in the rat, urine pH and fractional bicarbonate excretion were higher from the postobstructive kidney, while micropuncture measurements indicated no impairment of bicarbonate reabsorption in the proximal or distal tubules of surface nephrons (39). Walls et al (39) also found that tubular maximum (Tm) for bicarbonate studies showed no increased splay of the titration curves to suggest impaired reabsorption in the deep nephrons. These results, together with low urine P_{CO_2} values after bicarbonate loading, localize the defect in hydrogen-ion secretion (or bicarbonate reabsorption) to the terminal nephron (collecting duct, distal tubule) of the postobstructive kidney, although an abnormality in juxtamedullary nephron function cannot be excluded. The absence of defective acidification when postobstructive rats were acidemic (UUO given ammonium chloride, or BUO) suggests that decreased hydrogen-ion secretion in the terminal nephron (distal renal tubular acidosis) may only be seen when there is adequate bicarbonate delivery to this nephron segment.

Phosphate excretion by the postobstructive kidney is markedly decreased after relief of 24-hr UUO, despite an increase in the fractional excretion of sodium (40). The decrease in phosphate excretion was not present after relief of BUO when the filtered phosphate load was much higher and was not due to the tubules being refractory to parathyroid hormone. Decreased phosphate excretion by the unilateral postobstructive kidney appears to be due to decreased filtered load, with normal or enhanced proximal tubular reabsorption, while the dissociation from sodium excretion is due to inhibition of sodium reabsorption in the distal nephron.

URINE-CONCENTRATING DEFECT

The partially obstructed kidney, as well as the postobstructive kidney, are characterized by impairment of urinary-concentrating ability, which is out of proportion to the decrease in glomerular filtration rate. Nephrogenic diabetes insipidus has been well documented during partial obstruction (41) and POD is not responsive to vasopressin (6, 30). Unilateral ureteral obstruction for only 5 min with hydrostatic pressure of 100 mm Hg may cause reduced urine osmolality for several hours, in the absence of a persistent reduction in GFR and with no changes in electrolyte excretion (42). Urea and sodium concentration in medullary tissue are reduced and papillary water content is increased with brief periods of increased intrapelvic pressure (42, 43).

The concentrating defect does not appear to result from washout of medullary interstitial solute by continued blood flow during the period of obstruction, since clamping the renal artery does not prevent development of the defect (42, 43). Medullary washout of solute by increased blood flow in the immediate postobstructive period may be a significant factor but may not explain the defect in the partially obstructed kidney. The degree of increase in the intrapelvic pressure correlates with the concentrating defect (43), but how this effect is mediated remains unknown. Possible explanations include the removal of medullary solute by increased lymph flow, reduction in filtered sodium, or passive diffusion of solute out of the papilla into the pelvic urine. Sodium- and potassium-activated adenosine triphosphatase

(Na, K ATPase) is important in renal sodium reabsorption (44) and in concentrating capacity (45). Renal Na, K ATPase activity, which is highest in the outer medulla of the kidney (46), is decreased in the postobstructive kidney and correlates with changes in renal function (47). Finally, a striking abnormality in collecting-duct sodium and water transport has been demonstrated after relief of 24-hr complete obstruction in the rat, but the changes in the renal medulla that are responsible for these transport defects remain uncertain (18).

During chronic obstruction, total and individual-nephron GFRs are reduced (25), causing decreased delivery of solute to the medulla. A disproportionate decrease in GFR of juxtamedullary nephrons (19), and structural alterations, including shortening of the loops of Henle, vasa recta, and collecting ducts in the papilla, are undoubtedly also important in the impaired urine-concentrating ability.

CONCLUSION

The effects of urinary-tract obstruction on renal function have been clarified using modern physiologic techniques in animal models. These effects are obviously dependent on the severity and duration of obstruction, together with the presence or absence of extrarenal factors, particularly uremia, as exemplified by the differences in renal function between bilateral and unilateral obstruction. The beneficial effects on the unilateral postobstructive kidney of volume expansion (5) and the importance of azotemia in determining POD (20, 21) have possible clinical applications. Further experimental work is required to understand the mechanisms responsible for the altered renal hemodynamics and decreased GFR, as well as the abnormalities in medullary function resulting from obstruction. Improvement in the management of this common clinical problem may then be forthcoming.

ACKNOWLEDGMENTS

The author wishes to thank U. Honrath for expert technical assistance and C. Wright for secretarial help.

Literature Cited

1. Bricker, N. S., Klahr, S. 1971. *Obstructive Nephropathy In Diseases of the Kidney,* ed. M. B. Strauss, L. G. Welt, pp. 997–1037. Boston: Little, Brown
2. Yarger, W. E., Griffith, L. D. 1974. Intrarenal hemodynamics following chronic unilateral ureteral obstruction in the dog. *Am. J. Physiol.* 227:816–26
3. Moody, T. E., Vaughan, E. D. Gillenwater, J. Y. 1975. Relationship between renal blood flow and ureteral occlusion. *Invest. Urol.* 13:246–51
4. Selkurt, E. E., Deetjen, P., Brechtelsbauer, H. 1965. Tubular pressure gradients and filtration dynamics during urinary stop flow in the rat. *Pfluegers Arch.* 286:19–35
5. Jaenike, J. R. 1970. The renal response to ureteral obstruction: A model for the study of factors which influence glomerular filtration pressure. *J. Lab. Clin. Med.* 76:373–82
6. Yarger, W. E., Aynedjian, H. S., Bank, N. 1972. A micropuncture study of postobstructive diuresis in the rat. *J. Clin. Invest.* 51:625–37
7. Arendshorst, W. J., Finn, W. F., Gottschalk, C. W. 1974. Nephron stop-flow pressure response to obstruction for 24 hours in the rat kidney. *J. Clin. Invest.* 53:1497–1500
8. Suki, W., Eknoyan, G., Rector, F. C. Jr., Seldin, D. W. 1966. Patterns of nephron perfusion in acute and chronic

hydronephrosis. *J. Clin. Invest.* 45: 122–31

9. Fulop, M., Brazeau, P. 1970. Increased ureteral back pressure enhances renal tubular sodium reabsorption *J. Clin. Invest.* 49:2315–23
10. Lorentz, W. B. Jr., Lassiter, W. E., Gottschalk, C. W. 1972. Renal tubular permeability during increased intrarenal pressure. *J. Clin. Invest.* 51:484–92
11. Harris, R. H., Yarger, W. E. 1974. Renal function after release of unilateral ureteral obstruction in rats. *Am. J. Physiol.* 227:806–15
12. Buerkert, J., Purkerson, M. L., Klahr, S. 1976. On the site of decreased Na fluid reabsorption after release of ureteral obstruction in the rat. *J. Lab. Clin. Med.* 87:397–410
13. Feldman, R. A., Siegel, N. J., Kashgarian, M., Hayslett, J. P., Lytton, B. 1974. Intrarenal hemodynamics in postobstructive diuresis. *Invest. Urol.* 12:172–75
14. Wilson, B., Reisman, D. D., Moyer, C. A. 1951. Fluid balance in the urologic patient: Disturbance in the renal regulation of the excretion of water and sodium salts following decompression of the urinary bladder. *J. Urol.* 66:805–15
15. Bricker, N. S., Shwayri, E. I., Reardan, J. B., Kellog, D., Merrill, J. P., Holmes, J. H. 1957. An abnormality in renal function resulting from urinary tract obstruction. *Am. J. Med.* 23:554–64
16. Jaenike, J. R. 1972. The renal functional defect of postobstructive nephropathy: The effects of bilateral ureteral obstruction in the rat. *J. Clin. Invest.* 51:12, 2999–3006
17. McDougal, W. S., Wright, F. S. 1972. Defect in proximal and distal sodium transport in post-obstructive diuresis. *Kidney Int.* 2:304–17
18. Sonnenberg, H., Wilson, D. R. 1976. The role of the medullary collecting ducts in postobstructive diuresis. *J. Clin. Invest.* 57:1564–74
19. Wilson, D. R., Honrath, U. 1975. Nephron functional heterogeneity in the postobstructive kidney. *Kidney Int.* 7:19–26
20. Harris, R. H., Yarger, W. E. 1975. The pathogenesis of postobstructive diuresis: The role of circulating natriuretic and diuretic factors including urea. *J. Clin. Invest.* 56:880–87
21. Wilson, D. R., Honrath, U. 1976. Cross-circulation study of natriuretic factors in postobstructive diuresis. *J. Clin. Invest.* 57:380–89
22. Bourgoignie, J. J., Hwang, K. H., Ipakchi, E., Bricker, N. S. 1974. The presence of natriuretic factor in urine of patients with chronic uremia. The absence of the factor in nephrotic uremic patients. *J. Clin. Invest.* 53:1559–67
23. Bourgoignie, J. J., Hwang, K. H., Espinel, C., Klahr, S., Bricker, N. S. 1972. A natriuretic factor in the serum of patients with chronic uremia. *J. Clin. Invest.* 51:1514–27
24. Grantham, J. J., Irwin, R. L., Qualizza, P. B., Tucker, D. R., Whittier, F. C. 1973. Fluid secretion in isolated proximal straight renal tubules. Effect of human uremic serum. *J. Clin. Invest.* 52:2441–50
25. Wilson, D. R. 1972. Micropuncture study of chronic obstructive nephropathy before and after release of obstruction. *Kidney Int.* 2:119–30
26. Gillenwater, J. Y., Westervelt, F. B. Jr., Vaughan, E. D. Jr., Howards, S. S. 1975. Renal function after release of chronic unilateral hydronephrosis in man. *Kidney Int.* 7:179–86
27. Wilson, D. R. 1974. The influence of volume expansion on renal function after relief of chronic unilateral ureteral obstruction. *Kidney Int.* 5:402–10
28. Bercovitch, D. D., Kasen, L., Blann, L., Levitt, M. F. 1971. The postobstructive kidney: Observations on nephron function after the relief of 24 hr of ureteral ligation in the dog. *J. Clin. Invest.* 50:1154–65
29. Muldowney, F. P., Duffy, G. J., Kelly, D. G., Duff, F. A., Harrington, C., Freaney, R. 1966. Sodium diuresis after relief of obstructive uropathy. *N. Engl. J. Med.* 274:1294–1300
30. Falls, W. F. Jr., Stacy, W. F. 1973. Postobstructive diuresis: Studies in a dialyzed patient with a solitary kidney. *Am. J. Med.* 54:404–12
31. Wilson, D. R. 1975. Mechanisms of post-obstructive diuresis in the solitary hydronephrotic kidney of the rat. *Clin. Sci. Mol. Med.* 48:167–76
32. Kerr, W. S. Jr. 1954. Effect of complete ureteral obstruction for one week on kidney function. *J. Appl. Physiol.* 6:762–72
33. Vaughan, D. E. Jr., Sweet, R. E., Gillenwater, J. Y. 1973. Unilateral ureteral occlusion: Pattern of nephron repair and compensatory response. *J. Urol.* 109:979–82

34. Earlam, R. J. 1967. Recovery of renal function after prolonged ureteric obstruction. *Br. J. Urol.* 39:58
35. Brunschwig, A., Barber, H. R. K., Roberts, S. 1964. Return of renal function after varying periods of ureteral occlusion. *J. Am. Med. Assoc.* 188:125
36. Better, O. S., Arieff, A. I., Massry, S. G., Kleeman, C. R., Maxwell, M. H. 1973. Studies on renal function after relief of complete unilateral ureteral obstruction of three months' duration in man. *Am. J. Med.* 54:234–40
37. Ericsson, N. O., Wineberg, J., Zetterstrom, R. 1955. Renal function in infantile obstructive uropathy. *Acta Pediatr. Scand.* 44:444–59
39. Berlyne, G. M. 1961. Distal tubular function in chronic hydronephrosis. *Q. J. Med.* 30:339–55
39. Walls, J., Buerkert, J. E., Purkerson, M. L., Klahr, S. 1975. Nature of the acidifying defect after the relief of ureteral obstruction. *Kidney Int.* 7:304–16
40. Purkerson, M. L., Rolf, D. B., Chase, L. R., Slatopolsky, E., Klahr, S. 1974. Tubular reabsorption of phosphate after release of complete ureteral obstruction in the rat. *Kidney Int.* 5:326–36
41. Earley, L. E., 1956. Extreme polyuria in obstructive uropathy: Report of a case of "water-losing nephritis" in an infant, with a discussion of polyuria. *N. Engl. J. Med.* 255:600–5
42. Jaenike, J. R., Bray, G. A. 1960. Effects of acute transitory urinary obstruction in the dog. *Am. J. Physiol.* 199:1219–22
43. Honda, N., Aizawa, C., Morikawa, A., Yoshitoshi, Y. 1971. Effect of elevated ureteral pressure on renal medullary osmolal concentration in hydropenic rabbits. *Am. J. Physiol.* 221:698–703
44. Katz, A. I., Epstein, F. H. 1967. The role of sodium-potassium-activated adenosine triphosphatase in the reabsorption of sodium by the kidney. *J. Clin. Invest.* 46:1999–2011
45. Martinez-Maldonado M., Allen, J. C., Eknoyan, G., Suki, W. 1969. Renal concentrating mechanism: possible role for sodium-potassium activated adenosine triphosphatase. *Science* 165:807–8
46. Hendler, E., Torretti, J., Epstein, F. H. 1971. The distribution of sodium-potassium-activated adenosine triphosphatase in medulla and cortex of the kidney. *J. Clin. Invest.* 50:1329–37
47. Wilson, D. R., Knox, W., Hall, E., Sen, A. K. 1974. Renal sodium- and potassium-activated adenosine triphosphatase deficiency during post-obstructive diuresis in the rat. *Can. J. Physiol. Pharmacol.* 52:105–13

Ann. Rev. Med. 1977. 28:341–48
Copyright © 1977 by Annual Reviews Inc. All rights reserved

DEVELOPMENT OF CORONARY COLLATERAL CIRCULATION

❖7247

Andrew S. Wechsler, M.D.
Department of Surgery, Duke University Medical Center, Durham, North Carolina 27710

INTRODUCTION

Collateral coronary arteries provide alternate routes for delivery of blood to myocardium distal to major coronary-artery obstructions. These channels may be extracardiac, endomural, or arterioluminal in origin (1). One additional source of collateral myocardial blood flow is from deliberate attempts to surgically augment blood flow to ischemic regions of myocardium. Over the years these have included methods such as promotion of adhesion formation with inflammation and epicardial ingrowth, direct intramyocardial tunneling of bleeding arterial vessels, and, most recently, the construction of aortocoronary interposition saphenous-vein bypass grafts. This review deals primarily with intercoronary collateral vessels, since these constitute the major and best studied class of naturally occurring collateral arteries. These arteries have an anatomic point of origin termed the "stem," a mid-zone, and a reentry point. The presence, number, and transmural distribution of coronary collaterals appear to be determined genetically for different species and perhaps differences in collateral density within a given species may also be a hereditary characteristic (2, 3). For example, dogs form extensive epicardial collaterals, while pigs have primarily a dense subendocardial network of collateral arteries. In contrast, collateral vessels in man have been demonstrated at every depth of myocardium, but subendocardial connections appear most numerous. Goats, however, do not generally form collateral vessels (4, 5). Gradual closure of a coronary artery in the canine model rarely results in myocardial infarction, but similar preparations in pigs invariably result in some degree of tissue necrosis. The same technique applied to a goat usually results in the death of the animal from massive myocardial necrosis (6). Although man may form abundant collateral channels, it is unpredictable whether these prevent myocardial infarction even when the rate of occlusion is gradual. Although early anatomists demonstrated the presence of collateral arteries in man, considerable controversy surrounded these observations, with Schlesinger stating that interarterial anastomoses in normal hearts were not greater than 40 μm in diameter when there was no accompanying coronary-artery disease (7). However, Spalteholz perfused normal coronary arteries and demonstrated intercoronary anastomoses (8). Baroldi demonstrated these to exceed 100 μm using injection techniques (8), and Bloor demonstrated collateral channels up to 80 μm in diameter in hearts from infants (9).

Thorough and adequate assessment of coronary collateral development has proven difficult. Studies in one species may not supply information germane to other species, and the best experimental models mimic poorly human coronary-artery disease. Individual differences within the same species render difficult interpretation of studies done on different animals prepared in a similar manner, but studied at different times. Several studies have examined collateral development in the same animal. Such studies have generally been limited to either physiologic monitoring of hemodynamic effects of collateral development or to anatomic investigation of collateral evolution without hemodynamic assessment.

PHYSIOLOGIC MONITORING OF COLLATERAL DEVELOPMENT

Physiologic monitoring of collateral development is assessed by retrograde flow in the obstructed vessel, washout of radioactive tracers from affected regions of myocardium, measurements of peripheral coronary pressures, coronary arteriography, and the myocardial distribution of tracer microspheres. These parameters of collateral evolution are compared with indicators of diminishing native coronary flow such as may be determined by flow probes, pressure gradients, and washout rates of proximally injected inert radioactive gases (10–14). When such studies are performed, a pattern of coronary collateral function development appears and is typified by Figure 1 from an elegant paper by Elliot and his co-workers. An ameroid constrictor, flow transducer, occluding snare, and intracoronary-pressure catheter were all positioned on the circumflex coronary artery of dogs that were studied at frequent intervals. As can be seen in the figure, on about the 20th day, when blood flow in the circumflex coronary artery had fallen to approximately half the control level, and when reactive hyperemia had been almost abolished, mean peripheral circumflex pressure began rising. An aortic-circumflex pressure difference began increasing significantly. These observations indicated that collateral development began when peripheral vasodilatation was essentially maximal. As circumflex flow fell to zero over the ensuing 12 days, reactive hyperemia disappeared and the peripheral circumflex pressure rose to 70 mm Hg mean pressure and the aortic-circumflex pressure difference, which peaked at the 27th day, began to decline as collaterals provided more and more flow to the distal myocardium (14). Such observations have been repeated in a variety of other studies and generally yield

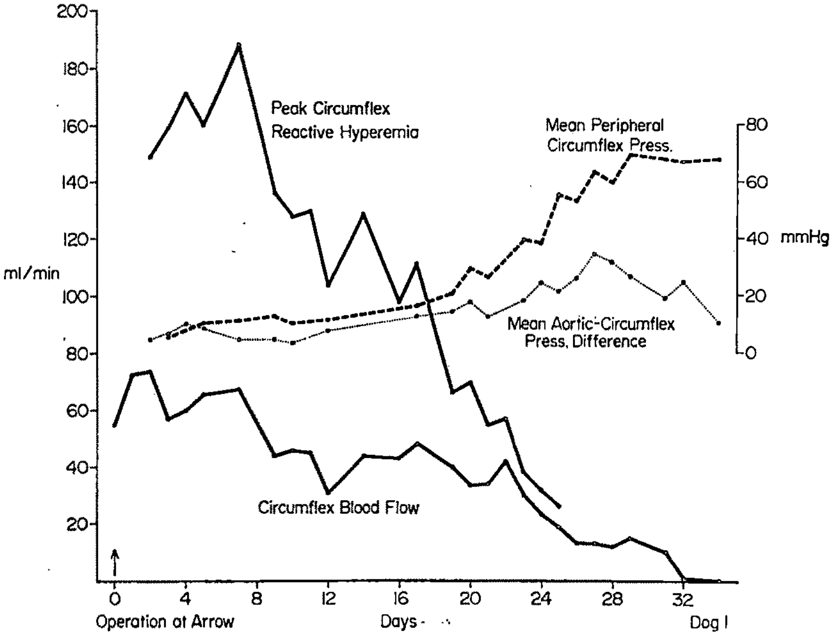


Figure 1 Sequential hemodynamic changes associated with gradual coronary occlusion in a dog.[1] Rise in peripheral coronary pressure occurred only after a pressure gradient existed across the circumflex obstruction despite significant circumflex flow reduction and distal coronary vasodilatation.

similar information (15). Moreover, the collateral blood supply at the end of this time is distributed across the myocardial wall in a pattern similar to normal distribution of coronary blood flow (16).

In the special circumstance of acute coronary-artery ligation, although infarction occurs, collateral development progresses at a rapid rate such that by 24 hr following the occlusion, 80% of normal blood-flow levels have been restored to the subepicardial region of the infarct zone as well as the periphery of this zone. The subendocardial region, however, retains only about 12% of its original flow without significant change in the first 24 hr following occlusion (17). In other studies, Khouri et al demonstrated that gradual occlusion of a coronary artery over a five-day period was associated with collateral development. When perfusion of the native coronary artery was reestablished for an hour and then reoccluded, within one hour physiologic evidence of collateral function was as apparent as prior to the allowance of reflow. These studies demonstrated clearly the persistence of potential collateral function and supported further the concept of major anatomic development (18).

[1] Published by permission of the American Heart Association, from (14).

ANATOMY OF CORONARY COLLATERAL DEVELOPMENT

Schaper examined in detail the anatomy of collateral-artery development (19). He demonstrated that development of coronary collaterals is an active process rather than simple passive stretching of preexisting channels and that this development could be divided into well-defined stages. The majority of developmental activity occurs in the mid-zone of preexisting collateral channels, this region generally representing the flow-limiting portion. Using ameroid constriction to produce gradual occlusion of the circumflex coronary artery, he examined multiple sections of collateral mid-zones at progressively increasing times from ameroid occlusion. The initial "transformation" stage of collateral development is acute dilatation and stretching in association with diminished flow through the native vessel. This dilatation and stretching is associated with rupture of the internal elastic lamina and thinning of the layer of smooth muscle cells. Perivascular inflammation occurs with blood-borne cells extruded from the capillaries into the surrounding myocardium occasionally accompanied by myocytolysis. Within one to two weeks after coronary occlusion, perivascular inflammation subsides and a proliferative stage begins with disorganized cell growth in the muscular wall layers and occasional mitoses in smooth muscle, epithelium, and fibroblasts. Six weeks after coronary occlusion, inflammatory changes diminish and smooth-muscle layering becomes more normal, regenerating an artery with both increased radius and wall thickness. Twelve months following coronary occlusion, vascular transformation and remodeling is complete, with the newly established collateral arteries resembling muscular arteries as much as 30 times the diameter of the original collateral from which they stemmed.

Tritiated thymidine administered to such preparations causes maximal labeling of nuclei eight weeks following ameroid implantation or about six weeks after closure. This is illustrated in Figure 2, taken from a study by Schaper. This correlates well with an abrupt slope change in the curve relating internal diameter of collaterals to time after ameroid implantation. Although further increments in vessel diameter occur during the next six months, these are of lesser magnitude than the multifold increase occurring in the first eight weeks.

STIMULI FOR COLLATERAL DEVELOPMENT

The preceding section documents that development of immature collateral vessels is initiated by dilatation and stretching of mid-zone portions of preexisting collateral channels. Vascular injury ensues, with vascular transformation occurring under a set of stimuli that favor large vessel development. The actual forces for collateral development are divided into mechanical, chemical, and hereditary categories. Although hereditary factors may condition the quality and quantity of collateral development, the exact nature of this influence is not defined. Chemical stimuli appear to be most important in the early phases initiating vascular dilatation. These may involve mechanisms similar to those operative during normal coronary vasodilatation associated with hypoxia or might depend upon release of substances

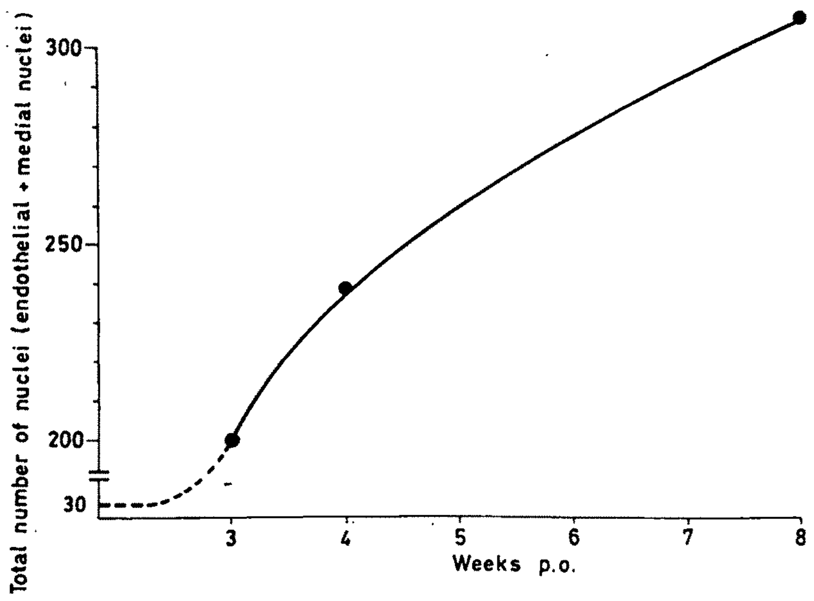


Figure 2 Tritiated thymidine–labeled nuclei following circumflex ameroid placement.[2] Maximal labeling occurs when organization of collateral vessels has followed a phase of disorganized proliferation.

from chronically ischemic regions of myocardium. The exact nature of these substances remains poorly defined, although mediators such as adenosine, hypoxia, Krebs-cycle intermediates, hyperosmolality, and potassium have all been implicated (20). The factors most easily discernible from physiologic monitoring are those related to altered physical forces in the capillary bed. Although flow velocity within collaterals, pressure gradients across the native coronary artery, and pressure inside the collateral vessels have all been suggested, the most likely factor is that of augmented wall stress and 'tangential shear' forces that are the consequence of vasodilatation, increased intraluminal pressure, and decreased wall thickness-to-radius ratios, all tending to increase tangential shear forces (21).

The extent to which collaterals actively alter their internal radius in response to metabolic and hemodynamic events has not been fully elucidated. It is likely that in the early stages of collateral formation the vessels act primarily as maximally dilated passive conduits (19). Much later in the course of collateral development, probably at the time when smooth-muscle walls have been reorganized, collaterals may respond actively to vasoconstrictor and vasodilator influences. The effect of various drugs, autonomic influences, and hormones on collateral formation are still under active investigation. Autonomic innervation of collaterals is scanty, with both fewer norepinephrine-containing granules in the vessel walls and evidence for vessel-wall separation from sympathetic nerve endings, possibly the consequence of perivascular inflammation present in the early stages of collateral development (11, 22). Some preliminary investigations suggest that growth hormone may augment early collateral formation, whereas hydrocortisone may diminish collateral formation, probably by interfering with the perivascular inflammatory response surrounding dilated vessels (9). Some evidence also exists that administration of nitrites, dipyridamole, or other strong vasodilators may increase collaterals; however, the complexity of relating findings in the literature with known stages of collateral development preclude a definitive conclusion at this time (19, 23, 24). Systemic hypoxia, anemia, and hypertrophy have all been associated with increased collateral formation (15, 25).

---

[2] Published by permission of the American Heart Association, from Schaper, W., DeBrandes, M., Lewi, P. 1971. DNA synthesis and mitoses in coronary collateral vessels of the dog. *Circ. Res.* 29:145.

COLLATERAL DEVELOPMENT IN MAN

Specific examinations of the collateral circulation and its development in man are fraught with difficulties. Arteriographic visualization of collaterals provides little insight into dynamics of collateral flow. Collateral flow estimates, as currently performed, are at best determined by indirect means. Pressure measurements obtained at operation are plagued by hazards in interpretation, due to patient variability, degrees of proximal stenosis, and changing conditions at the time of measurement. The primary unresolved question for man is whether or not the presence of a significant collateral circulation ensures protection of the myocardium from future ischemic events or whether such collateral development is primarily a marker of severe ischemic disease. Conflicting evidence is present as to whether patients with collaterals have greater or lesser degrees of functional capacity and longevity, as do patients with similar proximal obstructions but lesser collateral development. Clearly, collateral development increases progressively with increasing severity of coronary-artery occlusion, being greatest in patients with three-vessel disease and least in patients with single-vessel disease. In instances where aortocoronary saphenous-vein bypass grafts have been performed, and patients studied at later dates, data similar to that obtained by Elliot in dogs has been obtained regarding regression and reappearance of these collaterals. For patients with well-developed collateral circulations prior to bypass grafting, and with subsequent patent bypass grafts, these collaterals have not been demonstrated on subsequent angiography. On the other hand, in some of these same patients with subsequently occluded grafts, the collateral circulation has again been seen to be well visualized on arteriography (26, 27). Collateral development must be studied during periods of hemodynamic stress in order to fully assess its adequacy as a myocardial protective factor. In experimental animals with normal transmural coronary blood-flow distribution from collaterals to a region of obstructed circumflex coronary artery, infusion of norepinephrine resulted in significantly less increase of flow to regions of myocardium supplied by collaterals than to normal myocardium. When bypass grafts were placed distal to the ameroid obstruction, normal flow responses to norepinephrine infusion occurred (28). In parallel observations in man, when myocardial contractility was monitored, in the resting state no difference in function was seen when acutely placed bypass grafts were occluded or allowed to remain open. However, during norepinephrine infusion, marked increases in contractility occurred when bypass grafts were opened (29).

Future investigations must be concerned with methods for enhancing collateral formation, more attention to studies in subhuman primates, and the evolution of techniques allowing dynamic determination of collateral flow in man.

Literature Cited

1. Bloor, C. M., Liebow, A. A. 1965. Coronary collateral circulation. *Am. J. Cardiol.* 16:238–52
2. James, T. N. 1970. The delivery and distribution of coronary collateral circulation. *Chest* 58(3):183–203
3. Gregg, D. E., Fisher, J. C. 1963. Blood supply to the heart. In *Handbook of Physiology,* ed. W. F. Hamilton, Sect. 2, Vol. II. Washington DC: Am. Physiol. Soc. 1517 p.
4. Schaper, W., Jageneau, A., Xhonneux, R. 1967. The development of a collateral circulation in the pig and dog heart. *Cardiologia* 51:321
5. Schoenmackers, J. 1958. Zur Anatomie und Pathologie der Koronargenfasse. *Bad Oeynhausener Gesprache* 2:133
6. Rees, J. R. 1969. The myocardial collateral circulation. *Br. Heart J.* 31:1–4
7. Schlesinger, M. J. 1938. An injection plus dissection study of coronary artery occlusions and anastomoses. *Am. Heart J.* 15:528–68
8. Baroldi, G., Mantero, O., Scomazzoni, G. 1956. The collaterals of the coronary arteries in normal and pathologic hearts. *Circ. Res.* 4:223–29
9. Bloor, C. M., Keefe, J. F., Browne, M. J. 1966. Intercoronary anastomoses in congenital heart disease. *Circulation* 33:227–31
10. Mautz, F. R., Gregg, D. E. 1937. Dynamics of collateral circulation following chronic occlusion of coronary arteries. *Proc. Soc. Exp. Biol. Med.* 36:797–801
11. Gregg, D. E., Thornton, J. J., Mautz, F. R. 1939. Magnitude, adequacy and source of collateral blood flow and pressure in chronically occluded coronary arteries. *Am. J. Physiol.* 127:161–75
12. Kattus, A. A., Gregg, D. E. 1959. Some determinants of coronary collateral flow in the open-chest dog. *Circ. Res.* 7:628–42
13. Cibulski, A. A., Lehan, P. H., Timmis, H. H. 1972. Retrograde flow technique vs krypton-85 clearance technique for estimation of myocardial collaterals. *Am. J. Physiol.* 223:1081–87
14. Elliot, E. C., Jones, E. L., Bloor, C. M., Leon, A. S., Gregg, D. E. 1968. Day-to-day changes in coronary hemodynamics secondary to constriction of circumflex branch of left coronary artery in conscious dogs. *Circ. Res.* 227:237–50
15. Bloor, C. M. 1974. Functional significance of the coronary collateral circulation. *Am. J. Pathol.* 76(3):562–87
16. Becker, L. C., Pitt, B. 1971. Collateral blood flow in conscious dogs with chronic coronary artery occlusion. *Am. J. Physiol.* 221(5):1507–10
17. Cox, J. L., Pass, H. I., Wechsler, A. S., Oldham, H. N. Jr., Sabiston, D. C. Jr. 1975. Coronary collateral blood flow in acute myocardial infarction. *J. Thorac. Cardiovasc. Surg.* 69(1):117–25
18. Khouri, E. M., Gregg, D. E., McGranahan, G. M. Jr. 1971. Regression and reappearance of coronary collaterals. *Am. J. Physiol.* 220(3):655–61
19. Schaper, W. 1971. *The Collateral Circulation of the Heart.* Amsterdam: North-Holland
20. Haddy, F. J. Scott, J. B. 1968. Metabolically linked vasoactive chemicals in local regulation of blood flow. *Physiology* 48:688
21. Schaper, W. 1967 Tangential wall stress as a molding force in the development of collateral vessels in the canine heart. *Experientia* 23:595
22. Dresse, A., Xhonneux, R., Gerardy, J., Schaper, W. 1967. Innuence de l'hypoxie sur la teneur en catecholamines du myocarde de chien. *J. Physiol.* 59:395
23. Lumb, G. D., Hardy, L. B. 1963. Collateral circulation and survival related to gradual occlusion of the right coro-

nary artery in the pig. *Circulation* 27:717

24. Schmidt, H. D. Schmier, J. 1967. Erheblich veringerte Infarktmortalitat bei Hunden nach Vorbehandlung mit Pyrimidopyrimidion. *Arzneim.-Forsch.* 17:861
25. Eckstein, R. W. 1955. Development of interarterial coronary anastomoses by chronic anemia: disappearance following correction of anemia. *Circ. Res.* 3:306–10
26. Bourassa, M. G., Solignac, A., Goulet, C., Lesperance, J. 1974. Regression and appearance of coronary collaterals in humans during life. *Circulation* 49, 50:Suppl. II, pp. 127–35
27. Gregg, D. E. 1974. The natural history of coronary collateral development. *Circ. Res.* 35(3):335–44
28. Cox, J. L., Pass, H. I., Oldham, H. N. Jr., Wechsler, A. S., Sabiston, D. C. Jr. 1976. Coronary collateral circulation during stress and the effects of aorta-coronary bypass grafts. *J. Thorac. Cardiovasc. Surg.* 71(4):540–44
29. Wechsler, A. S., Sabiston, D. C. 1973. Effects of aorta-coronary bypass grafts on myocardial contractile state: response to catecholamine stress. In *Current Topics in Coronary Research,* ed. C. M. Bloor, R. A. Olsson, pp. 263–78. New York: Plenum

Ann. Rev. Med. 1977. 28:349–54
Copyright © 1977 by Annual Reviews Inc. All rights reserved

URATE EXCRETION

❖7248

George M. Fanelli, Jr., Ph.D.
The Merck Institute for Therapeutic Research, West Point, Pennsylvania 19486

INTRODUCTION

The renal handling of urate from both the physiologic and pharmacologic aspects, coupled with pathophysiologic implications, has generated new interest among some clinicians and basic scientists since the classic studies that emanated from the laboratory of the late Alexander B. Gutman. For very comprehensive reviews with reference to historical, comparative physiologic, pharmacologic, and therapeutic approaches, the reader should consult especially those of Emmerson (1), Gutman & Yü (2), Mudge, Berndt & Valtin (3), Steele & Rieselbach (4), Weiner & Fanelli (5, 6), Weiner & Mudge (7), Wesson (8), and Fanelli (9). Another review paper (10) deals exclusively with renal urate transport in nonhuman primates.

Gutman & Yü (11) in 1961 first put forth the now classic three-component system for urate excretion consisting of (*a*) complete ultrafilterability at the glomerulus, (*b*) subsequent tubular reabsorption, and (*c*) tubular secretion. Thus, uric acid appearing in the final urine represents a combination of the secreted fraction with that which may have escaped tubular reabsorption. Our current knowledge (5) of urate transport may be summarized as follows:

1. Most of urate reabsorption occurs in the proximal convoluted tubule.
2. Uricosuric agents act by depressing proximal reabsorption of urate.
3. Drugs that depress urate reabsorption act from the luminal (urine side) aspect of the tubule.
4. Urate is secreted in the proximal tubule.
5. The secretory flux for urate (and consequently the reabsorptive flux) is large.
6. A variety of compounds, uricosuric or not, inhibit the secretory flux.
7. The urate secretory mechanism is probably different from the *p*-aminohippurate mechanism.
8. Some agents that influence urate excretion are themselves secreted and reabsorbed by active transport.
9. A drug subject to bidirectional active transport may have bidirectional effects on urate transport.

This review does not include any aspects of gout or other disease states involving urate metabolism, as these conditions have been adequately dealt with elsewhere (12–15). It does, however, deal with uricosuric drugs, some of which are new, and other agents that influence urate excretion primarily in man.

URICOSURIC AGENTS

By definition, a uricosuric agent is one that increases the urinary excretion of uric acid. Clearly the most definitive treatment of this subject up to 1966 was that provided by Gutman (16). More recent but less exhaustive reviews (17, 18) have also appeared. Certain other agents such as salicylate and the antituberculous drug, pyrazinamide, or more correctly, its free acid (19) pyrazinoate (PZA) can induce urate retention or frank uricosuria when administered at low and high doses, respectively (20).

With salicylate this effect was first described by Yü & Gutman (21) and subsequently studied in detail by them (22). The use of salicylate for long-term therapy for gout is not feasible, due to numerous side effects noted with the high doses necessary for a uricosuric effect. For studies in man (23, 24), pyrazinamide is given, rather than the free acid, which is not an official preparation.

More than 25 years since its introduction, probenecid (Benemid®) continues to be of clinical utility (15) as a very effective uricosuric drug. Similarly, sulfinpyrazone (Anturan®), introduced in 1957, is a potent uricosuric agent for the treatment of gout. The latter two agents produce a similar incidence of gastric upset and not infrequently one is substituted for the other, depending upon individual requirements.

Two benzofuran derivatives, benziodarone and benzbromarone, have been developed in Europe. The former, used originally as a long-acting coronary vasodilator, was found to reduce serum urate and was therefore also uricosuric (25, 26). Benzbromarone, the newer analog, was reported to be a longer-acting uricosuric agent at relatively low oral doses in both normal and gouty subjects (27). More recently, Sorensen & Levinson (28), in an extensive study of benzbromarone lasting up to one year, reported that daily doses of 40–80 mg caused a rapid fall in plasma urate together with a concomitant marked increase in urate clearance. Benzbromarone also did not inhibit xanthine oxidase in this study, thus confirming earlier studies by Delbarre et al (25), and had no direct effect upon purine metabolic pathways but exerted its uricosuric action solely by inhibiting tubular reabsorption of urate.

The forementioned agents are all weak organic acids, but probably the most potent uricosuric drug was the weak base zoxazolamine (Flexin®) introduced in 1956 as a muscle relaxant. Burns and co-workers (29), while studying its metabolic disposition, fortuitously found it to be markedly uricosuric. They also noted that the principal metabolite, hydroxyzoxazolamine, was inactive as a uricosuric compound. Maroske & Weiner (30) conclusively indicated that zoxazolamine was not uricosuric in the dog. The lack of uricosuric activity in the dog is an example of fundamental differences in the urate transport systems that occur in the various species.

Species Differences Regarding Uricosuric Activity

Studies in this laboratory (31–33) directed specifically at uncovering why species differences are so common with respect to the magnitude and sensitivity to drugs of the renal urate-transport mechanisms have been performed employing a great-ape species, the chimpanzee. This species, like man, is devoid of the oxidative enzyme uricase, and its renal handling of urate is very similar to that of man. Also, zoxazolamine is extremely active as a uricosuric agent in the chimpanzee (32). It was ascertained that known uricosuric drugs in man were active in the chimpanzee and, more importantly, that the organomercurial diuretic mersalyl (Salyrgan®) readily produced net urate secretion, i.e. clearance ratios of urate-to-inulin (C_{urate}/C_{in}) were uniformly greater than unity (33). Efforts to duplicate the mersalyl results in man failed to show net urate secretion, but there was a marked uricosuric effect in that the maximal C_{urate}/C_{in} rose from a control value of less than 0.10 to approximately 0.50 (34). Taken in their entirety, it was concluded that the chimpanzee was a hyperresponder and man a hyporesponder to uricosuric drugs.

It should be remembered that up to 1975 all uricosuric agents were discovered in man only by chance.

New Agents Possessing Both Uricosuric and Diuretic Activity

A saluretic-uricosuric agent, (6, 7-dichloro-2-methyl-1-oxo-2-phenyl-5-indanyloxy)acetic acid, or MK-196, was found to be orally active at low doses in the chimpanzee by Fanelli et al (35). Probenecid did not appear to influence this dual action of MK-196. Mersalyl-induced net urate secretion was blocked by MK-196, indicating that MK-196, in addition to inhibiting the tubular reabsorption of urate, also limited its tubular secretory flux. The efficacy of this new agent was not compromised by changes in urinary pH. On a dosage-dependent basis, MK-196 was more saluretic (and uricosuric) than ethacrynic acid (Edecrin®) or furosemide (Lasix®) and possessed a longer duration of action.

Another uricosuric-diuretic compound [2, 3-dichloro-4-(2-thienylcarbonyl) phenoxy]acetic acid, known as SKF-62698 or ticrynafen, has been reported to be active in man, although at higher doses than are required for MK-196 (36, 37). It is also quite active as a uricosuric agent in the chimpanzee, but is only weakly natriuretic (unpublished observations). Reese & Steele (38) studied the effects of SKF-62698 on renal urate handling in normal subjects for two weeks in which plasma urate declined by more than 60%. Utilizing the pyrazinamide-suppression test, their results suggested that the excretion of secreted urate was enhanced after 2-week administration and reasoned that this was probably secondary to the inhibition of postsecretory urate reabsorption. SKF-62698 may well find a place in the treatment of diuretic-induced hyperuricemia but it may possibly, over longer periods of time, enhance the frequency of urolithiasis, since it produces very low levels of plasma urate. This would suggest that the proportion of total urate excreted in the urine increases as the plasma urate decreases. In other words, the fraction of daily urate production that is degraded in the intestine would decrease as the plasma concentration is reduced. Also, the potent uricosuric action of SKF-62698 in clinical doses could possibly leave a patient without adequate protection against excesses of dietary purines. Both factors would predispose to urolithiasis, not only the formation of uric acid stones, but also of calcium oxalate stones.

Other Uricosuric Agents in Man

Other agents that are uricosuric in man are halofenate (39–41) and the hypoglycemic sulfonylurea acetohexamide (or its metabolic product) reported by Yü, Burger & Gutman (42). The latter produced a modest uricosuria in gouty subjects with coexisting diabetes; acetohexamide was also uricosuric in the chimpanzee at comparable oral doses (unpublished data).

Various other drugs that are neither organic acids nor structurally related, such as outdated tetracycline (43), the tranquilizing agent chlorprothixene (44), and zoxazolamine, are uricosuric—possibly the result of some nonspecific disruption of urate reabsorption in the proximal tubule, although this tentative explanation may not apply to zoxazolamine.

Two anticholinergic agents, glycopyrrolate and tridihexethyl chloride, were noted to produce a slight increase in fractional uric-acid excretion in some hyperuricemic subjects but not in normouricemics (45). An expectorant, glyceryl guaicolate, had a mild uricosuric action in some gouty patients (46). It should also be noted that ethambutol, the antituberculous agent, caused hyperuricemia in some tuberculosis patients, together with a mean decrease in urate clearance of 22% relative to glomerular filtration rate. The mechanism of action of ethambutol seems to differ in one or more respects from pyrazinamide, diuretics, and ethanol, which also show a similar net reduction in urate clearance (47).

These nonspecific agents increase urate excretion by some as yet unknown means and are in all probability unrelated to the mechanisms of action of the organic acids, which have received considerably greater study. It must be stressed that none of the above agents are used as primary uricosuric drugs.

CONCLUSION

This review makes no attempt at being definitive; rather, a more general approach has been taken. Other reviews cited in the literature, taken together, present a more cogent picture of our current knowledge of all phases of renal urate excretion including historical background; physiology and pharmacology of urate transport in man, the nonhuman primates, and common laboratory animals; and current therapeutic regimens for gout and other pathologic states that alter urate excretion.

Many significant advances in our knowledge of urate excretion and uricosuric agents have been brought about by numerous workers over the years. It is hoped that similar advances and better therapeutic drugs will emerge in the years to come.

Literature Cited

1. Emmerson, B. T. 1974. Effects of drugs on the renal handling of uric acid. *Progr. Biochem. Pharmacol.* 9:142–52. Reprinted in *Drugs and the Kidney,* ed. K. D. G. Edwards, Basel: Karger. 274 pp.
2. Gutman, A. B., Yü, T. F. 1972. Renal mechanisms for regulation of uric acid excretion with special reference to normal and gouty man. *Semin. Arthritis Rheum.* 2:1–46
3. Mudge, G. H., Berndt, W. O., Valtin, H. 1973. Tubular transport of urea, glucose, phosphate, uric acid, sulfate, and thiosulfate. In *Handbook of Physiology,* ed. J. Orloff, R. W. Berliner, Sect. 8, pp. 587–652. Washington DC: Am. Physiol. Soc. 1082 pp.
4. Steele, T. H., Rieselbach, R. E. 1975. Renal urate excretion in normal man. *Nephron* 14:21–32
5. Weiner, I. M., Fanelli, G. M. Jr. 1974. Bidirectional transport: urate and other anions. In *Recent Advances in Renal Physiology and Pharmacology,* ed. L. G. Wesson, G. M. Fanelli Jr., pp. 53–68. Baltimore: Univ. Park Press. 388 pp.
6. Weiner, I. M., Fanelli, G. M. Jr. 1975. Renal urate excretion in animal models. *Nephron* 14:33–47
7. Weiner, I. M., Mudge, G. H. 1964. Renal tubular mechanisms for excretion of organic acids and bases. *Am. J. Med.* 12:743–62
8. Wesson, L. G. Jr. 1969. *Physiology of the Human Kidney,* p. 192. New York: Grune & Stratton. 712 pp.
9. Fanelli, G. M. Jr. 1976. Drugs affecting the renal handling of uric acid. In *Methods in Pharmacology,* ed. M. Martinez-Maldinado. New York: Plenum. 4A: 269–92
10. Fanelli, G. M. Jr., Beyer, K. H. Jr. 1974. Uric acid in nonhuman primates with special reference to its renal transport. *Ann. Rev. Pharmacol.* 14:355–64
11. Gutman, A. B., Yü, T. F. 1961. A three-component system for regulation of renal excretion of uric acid in man. *Trans. Assoc. Am. Physicians* 74:353–65
12. Becker, M. A., Seegmiller, J. E. 1974. Genetic aspects of gout. *Ann. Rev. Med.* 25:15–28
13. Wyngaarden, J. B., Kelley, W. N., Smith, L. H. Jr., Hughuley, C. M. Jr., Bain, J. A. 1972. In *The Metabolic Basis of Inherited Disease,* ed. J. B. Stanbury, J. B. Wyngaarden, D. S. Frederickson, pp. 889–1029. New York: McGraw-Hill
14. Newcombe, D. S. 1975. *Inherited Biochemical Disorders and Uric Acid Metabolism.* Baltimore: Univ. Park Press. 282 pp.
15. Proc. Conf. Gout Purine Metab. 2nd. 1975. *Arthritis Rheum.* 18:659–888 (Suppl.)
16. Gutman, A. B. 1966. Uricosuric drugs, with special reference to probenecid and sulfinpyrazone. *Adv. Pharmacol.* 4:91–141
17. Fanelli, G. M. Jr. 1975. Uricosuric agents. *Arthritis Rheum.* 18:853–58 (Suppl.)
18. Kelley, W. N. 1975. Pharmacologic approach to the maintenance of urate homeostatis. *Nephron* 14:99–115
19. Weiner, I. M., Tinker, J. P. 1972. Pharmacology of pyrazinamide: metabolic and renal function studies related to drug-induced urate retention. *J. Pharmacol. Exp. Ther.* 180:411–34
20. Fanelli, G. M. Jr., Weiner, I. M. 1973. Pyrazinoate excretion in the chimpanzee: relation to urate disposition and the actions of uricosuric drugs. *J. Clin. Invest.* 52:1946–57
21. Yü, T. F., Gutman, A. B. 1955. Paradoxical retention of uric acid by uricosuric drugs in low dosage. *Proc. Soc. Exp. Biol. Med.* 90:542–47
22. Yü, T. F., Gutman, A. B. 1959. Study of the paradoxical effects of salicylate in low, intermediate and high dosage on the renal mechanisms for excretion of urate in man. *J. Clin. Invest.* 38:1298–1315
23. Steele, T. H., Rieselbach, R. E. 1967. The renal mechanism for urate homeostasis in normal man. *Am. J. Med.* 43:868–75
24. Steele, T. H. 1973. Urate secretion in man: the pyrazinamide suppression test. *Ann. Intern. Med.* 79:734–37
25. Delbarre, F., Auscher, C., Olivier, J. L., Rose, A. 1967. Traitiment des hyperuricémies et de la goutte par des derivés du benzolyurane. *Sem. Hop.* 14:1127–33
26. Lemieux, G., Gougoux, A., Vinay, P., Michaud, G. 1973. Uricosuric effect of benziodarone in man and laboratory animals: a comparative study. *Am. J. Physiol.* 224:1431–39
27. Jain, A. K., Ryan, J. R., McMahon, F. G., Noveck, R. J. 1974. Effect of single oral doses of benzbromarone on serum

and urinary uric acid. *Arthritis Rheum.* 17:149–57

28. Sorensen, L. B., Levinson, D. J. 1976. Clinical evaluation of benzbromarone: a new uricosuric drug. *Arthritis Rheum.* 19:183–90
29. Burns, J. J., Yü, T. F., Berger, L., Gutman, A. B. 1958. Zoxazolamine: physiological disposition, uricosuric properties. *Am. J. Med.* 25:401–8
30. Maroske, D., Weiner, I. M. 1968. The renal handling of zoxazolamine (Flexin). *J. Pharmacol. Exp. Ther.* 159:409–15
31. Fanelli, G. M. Jr., Bohn, D. L., Reilly, S. S. 1971. Renal urate transport in the chimpanzee. *Am. J. Physiol.* 220: 613–20
32. Fanelli, G. M. Jr., Bohn, D. L., Reilly, S. S. 1971. Renal effects of uricosuric agents in the chimpanzee. *J. Pharmacol. Exp. Ther.* 177:591–99
33. Fanelli, G. M. Jr., Bohn, D. L., Reilly, S. S., Weiner, I. M. 1973. Effects of mercurial diuretics on renal transport of urate in the chimpanzee. *Am. J. Physiol.* 224:985–92
34. Fanelli, G. M., Hitzenberger, G. 1974. Uricosuric activity of intravenous mersalyl in man. *Medicon* 3:12–3
35. Fanelli, G. M. Jr., Bohn, D. L., Scriabine, A., Beyer, K. H. Jr. 1977. Saluretic and uricosuric effects of (6,7-dichloro-2-methyl-1-oxo-2-phenyl-5-indanyloxy)acetic acid (MK-196) in the chimpanzee. *J. Pharmacol. Exp. Ther.* 200:In press
36. Masbernard, A., Guadicelli, C. 1974. Étude clinique et de l'action antihypertensive de l'acide chloro-2, 3-(thienyl-2-ceto)-phenoxyace-tique en administration prolongée. *Lyon Med.* 232:165–74
37. Stote, R. M., Cherrill, D. A., Erb, B. B., Alexander, F. 1974. Site of action of SK&F 62698, a new diuretic-uricosuric agent. *Clin. Res.* 22:721A (Abstr.)
38. Reese, O. G., Steele, T. H. 1976. Renal transport of urate during diuretic-induced hypouricemia. *Am. J. Med.* 60:973–79
39. Jain, A., Ryan, J. R., Hague, D., McMahon, F. G. 1970. The effect of MK-185 on some aspects of uric acid metabolism. *Clin. Pharmacol. Ther.* 11:551–57
40. Morgan, J. P., Bianchine, J. R., Hsu, T.-H., Lasagna, L. 1971. Hypolipidemic, uricosuric, and thyroxine-displacing effects of MK-185 (Halofenate). *Clin. Pharmacol. Ther.* 12:517–24
41. Bluestone, R., Campion, D., Klinenberg, J. R. 1975. Halofenate: its selection and trial as a primary uricosuric agent. *Arthritis Rheum.* 18:859–62
42. Yü, T. F., Berger, L., Gutman, A. B. 1968. Hypoglycemic and uricosuric properties of acetohexamide and hydroxyhexamide. *Metab. Clin. Exp.* 17:309–16
43. Funlop, M., Drabkin, A. 1965. Potassium-depletion syndrome secondary to nephropathy apparently caused by "outdated tetracycline." *N. Engl. J. Med.* 272:986–89
44. Healey, L. A., Harrison, M., Decker, J. L. 1965. Uricosuric effect of chlorprothixene. *N. Engl. J. Med.* 272:526–27
45. Postelthwaite, A. E., Ramsdell, C. M., Kelley, W. N. 1974. Uricosuric effect of anticholinergic agent in hyperuricemic subjects. *Arch. Intern. Med.* 134:270–75
46. Ramsdell, C. M., Postelthwaite, A. E., Kelley, W. N. 1974. Uricosuric effect of glyceryl guaicolate. *J. Rheumatol.* 1:114–16
47. Postelthwaite, A. E., Bartel, A. G., Kelley, W. N. 1974. Hyperuricemia induced by ethambutol. In *Advances in Experimental Medicine and Biology, Purine Metabolism in Man,* ed. O. Sperling, A. De Vries, J. B. Wyngaarden, 41B:763–64. New York: Plenum. 846 pp.

Ann. Rev. Med. 1977. 28:355–69
Copyright © 1977 by Annual Reviews Inc. All rights reserved

THE EMERGENCE OF GROUP B STREPTOCOCCI IN INFECTIONS OF THE NEWBORN INFANT[1]

❖7249

Bascom F. Anthony, M.D. and Donald M. Okada, M.D.
Departments of Pediatrics and of Obstetrics and Gynecology, Harbor General Hospital Campus, UCLA School of Medicine, Torrance, California 90509

INTRODUCTION

Because of a long association with mastitis in dairy cattle, the group B streptococcus (*Streptococcus agalactiae*) is an agent of considerable veterinary and economic importance (1). However, it has achieved medical prominence only within the past 15 years. Since 1962, well over 500 instances of group B streptococcal sepsis[2] in newborn infants have been reported and many other cases, less stringently documented, have been described. In the United States, this organism is now the leading cause of serious bacterial infections in newborns (2–5) or at least shares this distinction with *Escherichia coli* (2). The incidence of group B streptococcal sepsis in several American and British centers has been estimated at between 1.3 and 3 per 1,000 live births (2, 6–9).

This chapter emphasizes clinical and relevant microbiological and epidemiological aspects of newborn streptococcal infections. The more general subject of human disease related to group B streptococci was reviewed by Eickhoff in 1972 (10).

HISTORICAL PERSPECTIVE

It is unknown whether the recent proliferation of reports of newborn streptococcal infections represents an actual increase in disease incidence and/or more accurate diagnosis. Although overshadowed as a neonatal pathogen in the earlier years by the group A streptococcus (11), group B was documented in several cases of sepsis between 1930 and 1960 (12, 13). In other reports compiled in the later years of this period, beta-hemolytic streptococci were isolated from septic infants but were not classified serologically (14, 15).

---

[1] The authors' studies cited here were supported by a research grant from the US Public Health Service (HD-07559).

[2] Unless indicated otherwise, the infant and adult infections considered here are confined to those in which bacteremia is documented by culture or can be presumed from the occurrence of hematogenous infection (e.g. meningitis, endocarditis, etc).

Group B streptococci have clearly been prevalent in humans for several decades, as components of the resident bacterial flora and as occasional agents of lethal or life-threatening infection. In 1935, Lancefield & Hare isolated them from the vaginal swabs of 4% of postpartum patients (16). It is striking that 18 cases of group B streptococcal sepsis in postpartum and postabortal females were reported in the decade prior to 1945 (12, 17–27), while a similar number have appeared in the years since (6, 28–35). Hence, one cannot easily document or explain the apparent recent increase in newborn streptococcal infections.

CHARACTERISTICS OF NEWBORN INFECTIONS

Although other authors had described group B streptococcal sepsis in the early hours and days of life (6) or several weeks later (36), the separate syndromes of "early" or "acute" infection and of "late" or "delayed" infection were first delineated in 1973 by Baker et al (37) and by Franciosi, Knostman & Zimmerman (7). Individual series have suggested a clear separation between the early and late syndromes by age of onset of symptoms, e.g. 48 hr (7) or 10 days (37), but Figure 1, which summarizes the data from 268 reported cases (4, 6, 34, 36–49; B. F. Anthony, D. M. Okada, unpublished observations), indicates a continuum. About 40% of all newborn streptococcal infections begin in the first 24 hr and about 70% in the first three days of life. There is a sharp fall in the number of cases early in the first week and a more gradual decline over the next three months. The group B streptocococci are rarely associated with serious disease in older infants and children (50, 51) and do not play a significant role in acute pharyngitis (52).

The principal features of the early and late syndromes are summarized in Table 1. In several series, early infection is associated with a significant incidence of premature birth (4, 7–9, 43, 46, 47, 53), rupture of membranes for over 24 hr before delivery (4–6, 8, 9, 53), and maternal fever and other signs of postpartum infection (6, 40, 46, 54). However, the early syndrome is also common in full-term infants in the absence of these obstetrical abnormalities. Conversely, although these abnormalities are not associated with late infection, this syndrome has been repeatedly observed in premature infants after several weeks of well-being and stability (34, 36, 55, 56).

In early and late infection, bacteremia is common and potentially fatal. A few infants, however, have apparently tolerated bacteremia without treatment for periods of 24 hr or more (2, 57). Many expressions of disease are of hematogenous origin, the most frequent being meningitis, which occurs in about 75% of infants with delayed disease and in approximately 30% of early infections (4, 6, 34, 36, 38–40, 42–46, 58; B. F. Anthony, D. M. Okada, unpublished observations). This resembles other forms of neonatal bacterial meningitis in clinical and spinal-fluid findings. However, group B streptococci have been isolated from cerebrospinal fluid in the absence of pleocytosis or necropsy evidence of meningitis (7, 46, 59). Although group B streptococci are more readily eradicated by antibiotic therapy from

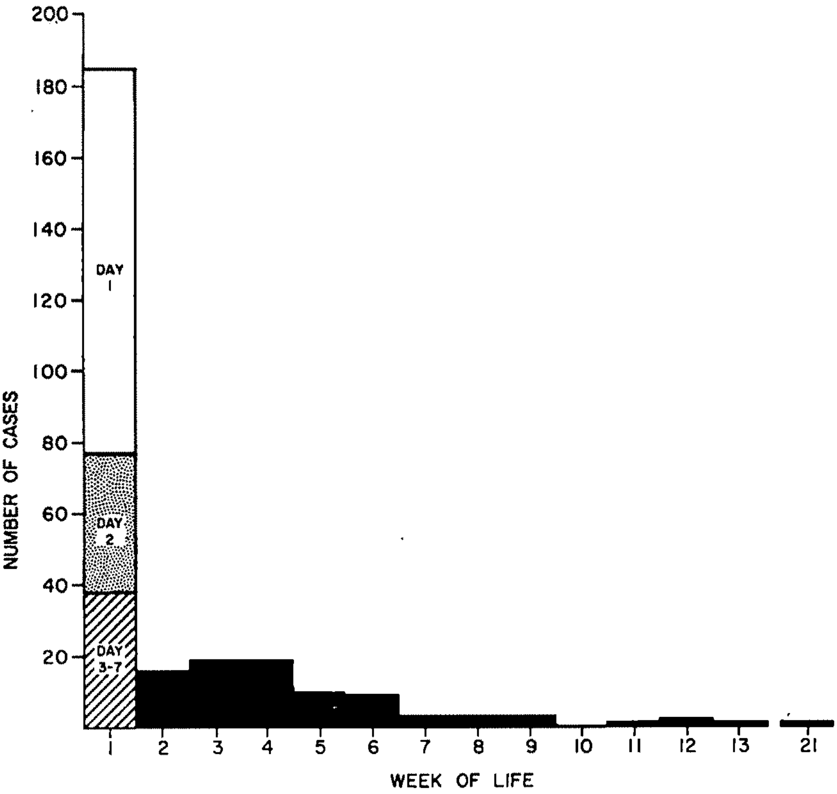


Figure 1 Age of first signs of group B streptococcal sepsis in 268 cases reported between 1962 and 1976, where the time of onset was specified. Series of three or less infants were excluded from this analysis.

the spinal fluid than *E. coli* and other gram-negative organisms (60), available data suggest that the prognosis of streptococcal meningitis may be as poor for immediate survival (2) and long-term sequelae (41, 59, 61). Diabetes insipidus has been a complication in several infants (62, 63; B. F. Anthony, D. M. Okada, unpublished observations).

Pulmonary disease is an interesting and poorly understood feature of streptococcal sepsis. Apparently rare in the delayed syndrome, it is diagnosed by X rays or necropsy in about 38% of early cases (4, 6, 34, 36, 38–40, 42–46, 53, 58; B. F. Anthony, D.M. Okada, unpublished observations). Acute respiratory distress is even more common as a clinical manifestation (4, 5, 7, 9, 37, 40, 43, 45). Histopathological descriptions of the pulmonary disease have not been uniform. Hyaline membranes have been prominent in several reports (46, 48, 53, 64), but not in one (7). The degree of pulmonary inflammatory response has ranged from insignificant (64) to moderate (7, 48, 53), and was variable in one series (46). A common finding has been the presence of large numbers of streptococci in the air spaces and interstitium of the lung (7, 48, 53, 64). The rapidly fatal course common in the pulmonary syndrome may be due in part to massive bacterial invasion from these sites. The clinical and radiological differentiation between streptococcal pulmonary disease and the respiratory distress syndrome of prematurity is very difficult, but certain clinical and radiological findings (48, 49, 53) and the development of respiratory distress in a relatively mature infant may provide clues in recognizing streptococcal infection.

Table 1 Characteristics of early and late newborn infections caused by group B streptococcus

	Early infection	Late infection
Incidence of prematurity	increased	not increased
Incidence of perinatal abnormalities[a]	increased	not increased
Common manifestations[b]	bacteremia meningitis (30%) pulmonary disease (38%)	bacteremia meningitis (75%)
Mortality rate[b]	55%	23%

[a] Prolonged rupture of membranes (>24 hours before birth) and maternal fever and other signs of infection.

[b] The incidence of meningitis and pulmonary disease and mortality rates were calculated from reports appearing between 1962 and 1976, which provided such information. Reports of three or less cases were excluded from this analysis, as were papers that specifically emphasized meningitis, pulmonary disease, or autopsy findings.

Other sites and manifestations of newborn streptococcal infection are remarkably varied. Hemorrhagic phenomena, possibly due to disseminated intravascular coagulation, have been described in several fatal infections (4, 43, 46). Pyarthrosis has been observed in a number of infants (2, 29, 34). Other unusual manifestations have included conjunctivitis (2), empyema (2, 38), endocarditis (39, 66), hematogenous endophthalmitis (G. R. Greene, personal communication), ethmoiditis (2), omphalitis (6, 67), osteomyelitis (2, 68), otitis media (69), pericarditis (70), peritonitis (66, 71), and both purulent and serous subdural effusions (62, B. F. Anthony, D. M. Okada, unpublished observations). Cutaneous lesions are rare, but have been attributed to the group B streptococci (2, 65).

The overall fatality rate for group B streptococcal sepsis in the newborn is 44%, based upon a compilation of 583 cases (2, 4, 5–7, 9, 29, 34, 36, 38–47, 53, 54, 58, 61, 72–74; B. F. Anthony, D. M. Okada, unpublished observations). It is significantly higher in early infection (55%) than in delayed sepsis (23%). The underlying condition of the infant has a strong influence on the mortality associated with streptococcal sepsis, which is particularly high in premature infants (46, 53).

MICROBIOLOGY

Diagnosis

Although definitive serological identification is available in reference laboratories and some hospitals, group B streptococci are commonly reported as "beta-hemolytic streptococcus, not group A." The distinction from group A is usually based upon a lack of inhibition by bacitracin discs and is not absolute, since 6–13% of group B strains are sensitive (44, 75). Other presumptive tests include the very reliable hippuricase-broth assay (75) and the CAMP reaction (76), which can be adapted for detecting nonhemolytic strains (77). The typical group B streptococcus is known to laboratory workers by its characteristic colony and narrow zone of weak hemolysis. However, some human isolates are completely nonhemolytic, a common finding with bovine strains (29, 78). In prenatal patients at Harbor General Hospital, 5–6% of group B strains are nonhemolytic (R. M. Rosenblatt, B. F. Anthony, D. M. Okada, unpublished observations). Nonhemolytic group B streptococci can be fully virulent and have been isolated from a number of septic infants (34, 79, 80).

Antibiotics

Group B streptococci resemble most other hemolytic streptococci in antibiotic profile (6, 34, 81, 82). To date, they have been uniformly sensitive to penicillin G, ampicillin, and other semisynthetic penicillins, but significantly higher penicillin levels are required against group B than group A streptococci. Since therapy of neonatal infections is commonly initiated with a penicillin and an aminoglycoside, it is of interest that group B streptococci are sensitive in vitro to the synergistic action of certain combinations and concentrations of these drugs (83, 84). It is unknown whether in vivo synergism occurs in the cerebrospinal and other body fluids of septic infants receiving antibiotic therapy.

Significance of Serotypes

The classification of most human isolates of group B streptococci into five serotypes (Ia, Ib, Ic, II, and III) is a valuable epidemiological tool and may be the basis for immunity in humans, as it is in laboratory animals. Specific associations between serotype and human virulence should be viewed cautiously, since all five types have been isolated from septic infants and since the prevalance of serotypes may shift with time and place. Type Ia was the most frequent isolate from a group of septic newborns in Holland (29) and types Ia, Ib, and Ic were most common in more recent reports from Colorado (7) and from India (47). It is of interest that the same distribution of serotypes was found in the maternal populations as in septic infants in two of the above reports (7, 29) and that type Ia was the most common serotype among a variety of human isolates at the Boston City Hospital 15 years ago (6). In the United States, type III is currently the leading serotype in both early and late newborn infections, type II is next, and types Ia, Ib, and Ic are least common (34, 58, 85). Since types II and III have been recently found to occur in similar proportions among healthy infants and their mothers (58, 86, 87), the data of Wilkinson, Facklam & Wortham (85) and Baker & Barrett (58) suggest that type III may indeed possess virulence out of proportion to its general prevalence. Figure 2 summarizes the experience at Harbor General Hospital. The serotype distribution from this small series of septic infants appears to reflect the serotypes isolated from the cords and mucous membranes of well newborns and from genital cultures of their mothers and is consistent with the possibility of increased virulence among type III strains.

Immunity

The nature of immunity against group B streptococci in experimental animals has been well defined, principally by Lancefield and associates. Serum from immunized rabbits is protective through the opsonization of streptococci and the activity is largely type-specific (88–90). Each of types Ia, Ib, and Ic contains either two or three antigens known to be shared among these serotypes (89, 91); antibody to any one may be opsonic and protective (89, 92).

Knowledge of human immunity is fragmentary. Although several laboratories have described opsonic activity in human serum against group B streptococci (90, 93, 94), further studies with precise, reproducible methodology are needed to define

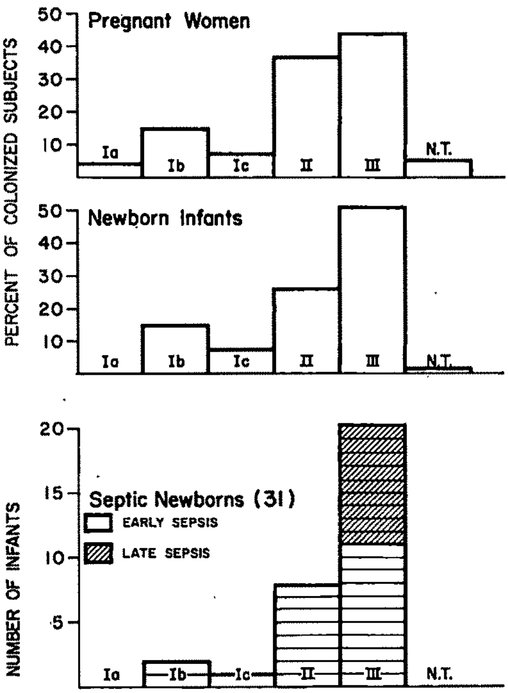


Figure 2 Distribution of group B serotypes isolated at Harbor General Hospital from healthy pregnant women (*top*) and their newborn infants (*middle*), and from a group of septic infants (*bottom*).

immunity in adults and infants. Important preliminary observations have been made by Baker and associates, who have demonstrated serum binding of an antigen with both type III and group B determinants in pregnant women who were colonized with type III, but whose newborn infants remained well; there was no binding by sera of infants with type III sepsis or of their mothers (95, 96). Using a chemiluminescence assay for phagocytosis, Hemming, Hall & Hill have reported a lack of opsonic activity against the infecting serotype in the sera of septic newborns and their mothers (97). These studies suggest that immunological methods may be used to identify mothers and infants at increased risk of newborn streptococcal infection.

EPIDEMIOLOGY

Adults

The principal human reservoir of the group B streptococcus is considered to be the female genital tract. Group B streptococci have been isolated from 4–25% of pregnant and parturient women cultured on one occasion (7–9, 16, 44, 86, 87, 98–105). The proportion of women identified as carriers can be doubled with repeated cultures (86–87). Rates of isolation from throat culture in the same subjects have been substantially lower (1–5%) (7, 8, 86, 106). Isolation rates from the female genital tract are significantly higher with selective broth than with direct inoculation of blood agar (103, 106, 107) and higher from vulvar and anterior vaginal swabs than from direct cervical swabs, in both the pregnant (44, 86, 99, 104, 108) and nonpregnant states (107, 109, 110). Group B streptococci are said to be more frequent in rectal than in vaginal cultures of pregnant women (111); it is unknown whether isolations from the rectum reflect transfer from the perineum or the gastrointestinal flora. Group B streptococci were not identified in the stools of pregnant women in one earlier report (112), but were recovered from 5.5% of human fecal samples in another (113).

Many reported differences in prevalence rates can be attributed to variables in culture methods, as described above. As suggested in a recent study (103), there may be true variations in prevalence between different communities and populations. However, relationships between streptococcal prevalence and age, parity, hygiene, birth-control practices, and socioeconomic status are largely unexplored. Suggestions of a lower incidence of group B streptococcal sepsis in black infants (4, 37, 43), but of higher rates of vaginal streptococci in black pregnant women (87) warrant further evaluation in a single population. With the exception of one report (106), there appear to be no significant variations among different stages of pregnancy, labor, or the postpartum period (7, 86, 87; C. J. Baker, personal communication). Strictly comparative data on pregnant and nonpregnant females are not available, but group B streptococci appear to be common in both states (7, 8, 87, 103, 107, 109, 110, 114).

Most data on carriage of group B streptococci in the birth canal are prevalence rates based upon one point in time and there is unfortunately little information from repeated or long-range observations. The authors have identified some women in whom the same group B serotype is consistently isolated throughout pregnancy and others in whom carriage appears to be transient (86). Bergqvist and associates have isolated group B streptococci from a small group of women four years after they were identified as carriers during pregnancy (114).

Group B streptococci have been isolated at widely varying rates (2.5–45%) from urethral cultures of adult males (7, 103, 107, 109). These isolations suggest that venereal transmission may be significant. However, Wallin & Forsgren could demonstrate no relationship between the genital carriage of group B streptococci and signs of infection or sexual promiscuity in male or female patients attending a venereal-disease clinic (107). Similarly, Baker and colleagues showed that prevalence of group B streptococci in a group of college women was unrelated to history of sexual activity (115).

Infants

The onset of symptoms in the first hours of life, the prevalence of group B streptococci in the birth canal, the postpartum isolation of these bacteria from the mothers of septic infants (4, 40, 43, 44) and the identity of maternal and infant serotypes (7) point to a maternal source for most early infections. The epidemiology of delayed infections may be more complex. Positive postpartum cultures in a number of late cases (7, 40, 43, 44) have been consistent with transmission from the mother. Negative vaginal cultures at the time of the infant's illness (4, 7, 116) do not exclude a maternal origin, since spontaneous loss of streptococci during pregnancy or after delivery is not unusual (86, 104). The occurrence of newborn infections in clusters of two or three has suggested nosocomial spread of group B streptococci (116–118), a possibility supported by considerable indirect evidence in one such outbreak (116). However, any connection among cases was effectively ruled out by serotyping in another (119). Aber and his colleagues have documented a case of delayed sepsis in which prospective observations appear to have excluded transmission from the mother (87) and several studies leave little doubt that *healthy* infants may acquire group B streptococci from other sources.

Simultaneous studies of infant-mother populations indicate similar prevalence rates (8, 29, 87, 98). Higher rates are found when multiple sites in the infant are examined. Group B streptococci are more commonly isolated from cord, external ear, and rectum than from throat swabs (7, 8, 111). Prospective observations have demonstrated that when group B streptococci are isolated from the birth canal at delivery, they may be acquired by 58–72% of the infants (8, 86, 87, 101). Conversely, a substantial but smaller proportion (10–25%) of newborns of culture-negative mothers acquire group B streptococci in the first week of life (8, 86, 87) and infants of carrier mothers sometimes acquire a different serotype of group B (B. F. Anthony, D. M. Okada, unpublished observations). Consequently, almost one half of neonatal acquisitions in the first week appear unrelated to the mother (86, 87) and are presumed to be nosocomial in origin. The frequency with which young infants may acquire group B streptococci in the home environment is unknown.

Newborn streptococcal acquisitions from the mother are presumed to occur usually during the birth process. An important exception is infection in utero after the membranes have been ruptured for a significant period; an inverse relationship has been demonstrated between the interval from membrane rupture to delivery and the age of onset of symptoms of early sepsis (4). Group B streptococcal infections have also been documented in infants delivered by cesarian section with visibly intact membranes (6, 74, B. F. Anthony, D. M. Okada, unpublished observations).

Antibiotic Prophylaxis

Antibiotic treatment of the mother has been proposed, with the goal of eradicating group B streptococci from the birth canal, thereby preventing colonization and sepsis in the newborn. Preliminary, uncontrolled observations in a group of first- and second-trimester pregnancies were encouraging (7), but the recent observations of Hall and his colleagues indicate limitations to this approach. One half of approximately 60 women identified as carriers in the third trimester were given ampicillin orally for seven days; three weeks later, the number with demonstrable group B streptococci was significantly less than among untreated subjects. At delivery, however, the proportions of colonized women and infants were similar in treatment and control groups (104). Thus, antibiotic therapy may eradicate or suppress group B streptococci in the birth canal, but the effect is temporary and the streptococcus can often be expected to reappear. Data on the antibiotic eradication of group B streptococci from superficial body sites in the newborn are more limited, but suggest that intramuscular benzathine penicillin and oral penicillin V are relatively ineffective (116) and that even more intensive penicillin therapy in infants (receiving successful treatment for sepsis) does not consistently eliminate streptococci from the throat, umbilicus, and rectum (120). Whether these penicillin "failures" in women and infants represent regrowth of persisting streptococci or reinfection from an exogenous source, they may be relevant to the unusual second attacks of streptococcal sepsis described in infants who have responded to therapy for an earlier infection (4, 121–123).

PROSPECTS FOR PREVENTION

The extent and severity of newborn infections with group B streptococci are well documented and widely recognized. The survival rates cited here were observed in the last 15 years and there is little prospect for significant improvement with current or anticipated methods of treatment. Unfortunately, the formulation of effective preventive measures is limited by our ignorance of the natural history, epidemiology, and immunology of group B streptococcal colonization and disease.

The tacit simplicity of a streptococcal culture and the demonstrated significance of maternal streptococci in the more severe and more common early syndrome are formidable arguments for routine prenatal screening for group B streptococci. However, a number of serious, unresolved questions prevent standardization and evaluation of this approach. These include the determination of the optimal time(s) in pregnancy for surveillance, the selection of maternal body site(s) to be sampled, the choice of laboratory techniques that are appropriately sensitive and widely practical, the reliability of antepartum cultures in predicting the presence or absence of group

B streptococci at delivery, and, most significantly, the establishment of reasonable and effective management of the carrier-mother and her infant.

The limitations of antibiotic prophylaxis that have been described to date should not preclude additional trials of penicillin and possibly of other drugs. Evaluation of prophylaxis in the adult female must reckon with the questions of venereal spread, of the male as an infectious reservoir of the organism, and of whether treatment of both sexual partners may effectively eradicate the group B streptococcus. Antibiotic therapy of the mother during labor or of the infant from the moment of birth, before bacterial growth and colonization, may conceivably interrupt mother-to-infant transmission, but such measures have not been carefully evaluated.

On the basis of comparisons between rates of disease incidence and of group B streptococcal colonization within newborn populations, it has been estimated that for every infant who develops streptococcal sepsis, approximately 100 colonized infants can be identified (8, B. F. Anthony, D. M. Okada, unpublished observations). Thus, if a chemoprophylactic regimen were proven to be effective, the present state of understanding would require its application to large numbers of colonized but otherwise healthy women and/or their infants.

Problems that can be anticipated in the development of a bacterial vaccine for administration during pregnancy and the childbearing years, as well as uncertainties in the selection of antigen(s) and in the prediction of prevalent serotypes, suggest that the prospects for active immunization against group B streptococci are more remote, but additional studies are needed.

The group B streptococci are uniquely and inexplicably virulent for infants in the perinatal period of life. The characteristic age distribution of group B streptococcal disease (Figure 1) is in contrast to the older age distribution of infections due to *Hemophilus influenzae* type B and *Neisseria meningitidis,* in which the presence or absence of protective antibody appears to be the principal determinant of disease (124, 125). It is possible that immunity is one of several determinants of group B streptococcal sepsis and that immunological studies can be used to identify pregnancies with an increased risk of newborn infection. Much more needs to be learned about the origins, prevalence, protective mechanisms, duration, placental transfer, and particularly the relationships of immunity to maternal colonization and newborn infection; prenatal testing for group B streptococcal immunity might then be used to select susceptible pregnant women and their infants for special monitoring and preventive measures.

Literature Cited

1. Schalm, O. W., Carrol, E. J., Jain, N. C., eds. 1971. *Bovine Mastitis.* Philadelphia: Lea & Febiger. 360 pp.
2. Howard, J. B., McCracken, G. H. Jr. 1974. The spectrum of group B streptococcal infections in infancy. *Am. J. Dis. Child.* 128:815–18
3. Yow, M. D. 1975. In *Infections of the Fetus and the Newborn Infant,* ed. S. Krugman, A. A. Gershon, 3:159–66. New York: Liss. 193 pp.
4. Tseng, P., Kandall, S. R. 1974. Group B streptococcal disease. *NY State J. Med.* 74:2169–73
5. Farquhar, J. Jr., Andrews, B. F. 1975. The changing bacteriologic pattern of newborn septicemia. *J. Ky. Med. Assoc.* 73:657–59
6. Eickhoff, T. C., Klein, J. O., Daly, A. K., Ingall, D., Finland, M. 1964. Neonatal sepsis and other infections due to group B beta-hemolytic streptococci. *N. Engl. J. Med.* 271:1221–28
7. Franciosi, R. A., Knostman, J. D., Zimmerman, R. A. 1973. Group B streptococcal neonatal and infant infections. *J. Pediatr.* 82:707–18
8. Baker, C. J., Barrett, F. F. 1973. Transmission of group B streptococci among parturient women and their neonates. *J. Pediatr.* 83:919–25
9. Reid, T. M. S. 1975. Emergence of group B streptococci in obstetric and perinatal infections. *Br. Med. J.* 2: 533–35
10. Eickhoff, T. C. 1972. In *Streptococci and Streptococcal Diseases,* ed. L. W. Wannamaker, J. M. Matsen, 533–43. New York: Academic. 635 pp.
11. Dunham, E. C. 1933. Septicemia in the new-born. *Am. J. Dis. Child.* 45:229–53
12. Brown, J. H. 1939. Double-zone beta-hemolytic streptococci. *J. Bacteriol.* 37:133–44
13. Nyhan, W. L., Fousek, M. D. 1958. Septicemia of the newborn. *Pediatrics* 22:268–78
14. Ziai, M., Haggerty, R. J. 1958. Neonatal meningitis. *N. Engl. J. Med.* 259:314–20
15. Groover, R. V., Sutherland, J. M., Landing, B. H. 1961. Purulent meningitis of newborn infants. *N. Engl. J. Med.* 264:1115–22
16. Lancefield, R. C., Hare, R. 1935. The serological differentiation of pathogenic and non-pathogenic strains of hemolytic streptococci from parturient women. *J. Exp. Med.* 61:335–49
17. Congdon, P. M., Manch, M. D. 1935. Streptococcal infection in childbirth and septic abortion. *Lancet* 2:1287–88
18. Colebrook, L., Purdie, A. W. 1937. Treatment of 106 cases of puerperal fever by sulphanilamide. *Lancet* 2: 1237–42
19. Fry, R. M. 1938. Fatal infections by haemolytic streptococcus group B. *Lancet* 1:199–201
20. Hill, A. M., Butler, H. M. 1940. Haemolytic streptococcal infections following childbirth and abortion. *Med. J. Aust.* 1:293–99
21. Pomales-Lebrón, A. 1940. A study of hemolytic streptococci as found in the tropical island of Puerto Rico. *PR J. Public Health Trop. Med.* 16:66–199
22. Rosenthal, A. H., Stone, F. M. 1940. Puerperal infection with vegetative endocarditis. *J. Am. Med. Assoc.* 114:840–43
23. Koletsky, S. 1941. Case of acute bacterial endocarditis and septicemia during the puerperium. *Ohio State Med. J.* 37:866–67
24. Ramsay, A. M., Gillespie, M. 1941. Puerperal infection associated with haemolytic streptococci other than Lancefield's group. A. *J. Obstet. Gynaecol. Br. Commonw.* 48:569–85
25. Rantz, L. A., Keefer, C. S. 1941. The distribution of hemolytic streptococci groups A, B, and C in human infections. *J. Infect. Dis.* 68:128–32
26. Gaustad, V. 1942. En oversikt over de forskjellige serologiske streptokokkgruppers betydning for menneskepatologien, belyst ved 2 tilfelle av streptokokksepsis fremkalt av hemolytiske streptokokker tilhørende henholdsvis Lancefields gruppe B og H. *Nord. Med.* 15:2026–29
27. Rantz, L. A., Kirby, W. M. M. 1942. Hemolytic streptococcus bacteremia. *N. Engl. J. Med.* 227:730–33
28. Dolphin, A., Cruickshank, R. 1945. Penicillin therapy in acute bacterial endocarditis. *Br. Med. J.* 1:897–901
29. Butter, M. N. W., DeMoor, C. E. 1967. *Streptococcus agalactiae* as a cause of meningitis in the newborn, and of bacteraemia in adults. *Antonie van Leeuwenhoek; J. Microbiol. Serol.* 33:439–50
30. Toft, H., Jespersen, C. S. 1968. Meningitis forarsaget af gruppe B β-haemolytiske streptokokker. *Nord. Med.* 17:1401–2

31. Wolstan, B. J. 1973. Group B streptococcal meningitis in adults *Minn. Med.* 56:631–32
32. Grossman, J., Tompkins, R. L. 1974. Group B beta-hemolytic streptococcal meningitis in mother and infant. *N. Engl. J. Med.* 290:387–88
33. Vilde, J. L., Roujeau, J. C., Bure, A., Verliac, F., Bastin, R. 1974. Infections graves et septicemies a streptocoque du groupe B chez l'adulte. *Sem. Hop.* 50:355–62
34. Anthony, B. F., Concepcion, N. F. 1975. Group B streptococcus in a general hospital. *J. Infect. Dis.* 132:561–67
35. Bayer, A. S., Chow, A. W., Anthony, B. F., Guze, L. B. 1976. Serious adult infections due to group B streptococci: Clinical and serotypic characterization. *Am. J. Med.* 61:498–503
36. Reinarz, J. A., Sanford, J. P. 1965. Human infections caused by non-group A or D streptococci. *Medicine* 44:81–96
37. Baker, C. J., Barrett, F. F., Gordon, R. C., Yow, M. D. 1973. Suppurative meningitis due to streptococci of Lancefield group B: A study of 33 infants. *J. Pediatr.* 82:724–29
38. Bouar, Y. L., Trung, P. H., Mozziconacci, P. 1970. Les meningites neonatales a streptocoque du groupe B. *Ann. Pediatr.* 17:207–13
39. Sander, C., Steinitz, H., Lösel, I. 1970. Meningitis und sepsis im neugeborenenalter durch B-streptokokken (*streptococcus agalactiae*). *Monatsschr. Kinderheilkd.* 118:640–43
40. Bergqvist, G., Hurvell, B., Balmborg, A. S., Rylander, M., Tunell, R. 1971. Neonatal infections caused by group B streptococci. *Scand. J. Infect. Dis.* 3:157–62
41. Cayeux, P. 1972. Infections neonatales a streptocoques du groupe B constatations etiologiques. *Arch. Fr. Pediatr.* 29:391–96
42. Echeverria, P. 1973. Observations concerning infections with beta hemolytic streptococci, not group A or D, in neonates. *J. Pediatr.* 83:499–511
43. Hey, D. J., Hall, R. T., Burry, V. F., Thurn, A. N. 1973. Neonatal infections caused by group B streptococci. *Am. J. Obstet. Gynecol.* 116:43–47
44. Bevanger, L. 1974. Carrier rate of group B streptococci with relevance to neonatal infections. *Infection* 2:123–26
45. Horn, K. A., Meyer, W. T., Wyrick, B. C., Zimmerman, R. A. 1974. Group B streptococcal neonatal infection. *J. Am. Med. Assoc.* 230:1165–67
46. Quirante, J., Ceballos, R., Cassady, G. 1974. Group B β-hemolytic streptococcal infection in the newborn. *Am. J. Dis. Child.* 128:659–65
47. Sharma, K. B., Prakash, K., Ravindran, P. C., Narang, P. 1975. Neonatal septicemia due to group B-beta hemolytic streptococci. *Indian J. Med. Res.* 63:781–86
48. Ablow, R. C., Driscoll, S. G., Effmann, E. L., Gross, I., Jolles, C. J., Uauy, R., Warshaw, J. B. 1976. A comparison of early-onset group B streptococcal neonatal infection and the respiratory-distress syndrome of the newborn. *N. Engl. J. Med.* 294:65–70
49. Weller, M. H., Katzenstein, A. A. 1976. Radiological findings in group B streptococcal sepsis. *Radiology* 118:385–87
50. Wheeler, S. M., Foley, G. E. 1943. A note on non-group-A streptococci associated with human infection. *J. Bacteriol.* 46:391–92
51. Sicherman, M. J. 1965. Septicemia due to group B streptococcus: Report of two siblings presenting with lesions resembling erythema nodosum. *Pediatrics* 36:937–39
52. The Commission on Acute Respiratory Diseases. 1947. The role of Lancefield groups of beta-hemolytic streptococci in respiratory infections. *N. Engl. J. Med.* 236:157–66
53. Vollman, J. H., Smith, W. L., Ballard, E. T., Light, I. J. 1976. Early onset group B streptococcal disease: Clinical, roentgenographic, and pathologic features. *J. Pediatr.* 89:199–203
54. Bobitt, J. R., Ledger, W. J. 1976. Obstetric observations in eleven cases of neonatal sepsis due to the group B β hemolytic streptococcus. *Obstet. Gynecol.* 47:439–42
55. Maher, E., Irwin, R. C. 1966. Group B streptococcal infection in infancy: A case report and review. *Pediatrics* 38:659–60
56. Jones, H. E., Howells, C. H. L. 1968. Neonatal meningitis due to *Streptococcus agalactiae. Postgrad. Med. J.* 44:549–62
57. Roberts, K. B. 1976. Persistent group B streptococcus bacteremia without clinical "sepsis" in infants. *J. Pediatr.* 88:1059–60
58. Baker, C. J., Barrett, F. F. 1974. Group B streptococcal infections in infants. *J. Am. Med. Assoc.* 230:1158–60
59. Horn, K. A., Zimmerman, R. A., Knostman, J. D., Meyer, W. T. 1974. Neurological sequelae of group B strep-

tococcal neonatal infection. *Pediatrics* 53:501–4

60. McCracken, G. H. Jr. 1972. The rate of bacteriologic response to antimicrobial therapy in neonatal meningitis. *Am. J. Dis. Child.* 123:547–53
61. Barton, L. L., Feigin, R. D., Lins, R. 1973. Group B beta hemolytic streptococcal meningitis in infants. *J. Pediatr.* 82:719–23
62. McReynolds, E. W., Roy, S. 1974. Diabetes insipidus secondary to group B beta streptococcal meningitis. *J. Tenn. Med. Assoc.* 67:117–20
63. Pai, K. G., Rubin, H. M., Wedemeyer, P. P., Linarelli, L. G. 1976. Hypothalamic-pituitary dysfunction following group B beta hemolytic streptococcal meningitis in a neonate. *J. Pediatr.* 88:289–91
64. Katzenstein, A. L., Davis, C., Braude, A. 1976. Pulmonary changes in neonatal sepsis due to group B β-hemolytic streptococcus. Relation to hyaline membrane disease. *J. Infect. Dis.* 133:430–35
65. Belgaumkar, T. K. 1975. Impetigo neonatorum congenita due to group B beta-hemolytic streptococcus infection. *J. Pediatr.* 86:982–83
66. Chattopadhyay, B. 1975. Fatal neonatal meningitis due to group B streptococci. *Postgrad. Med. J.* 51:240–43
67. VanPeenen, P. F. D., Cannon, R. E., Seibert, D. J. 1965. Group B beta-hemolytic streptococci causing fatal meningitis in an infant. *Mil. Med.* 130:65–67
68. Siskind, B., Galliguez, P., Wald, E. R. 1975. Group B beta hemolytic streptococcal osteomyelitis/purulent arthritis in neonates: Report of three cases. *J. Pediatr.* 87:659–62
65. Ermocilla, R., Cassady, G., Ceballos, R. 1974. Otitis media in the pathogenesis of neonatal meningitis with group B beta-hemolytic streptococcus. *Pediatrics* 54:643–44
70. Harper, I. A. 1971. The importance of group B streptococci as human pathogens in the British Isles. *J. Clin. Pathol.* 24:438–41
71. Hill, H. R., Riter, M. E., Menge, S. K., Johnson, D. R., Matsen, J. M. 1975. Rapid identification of group B streptococci by counterimmunoelectrophoresis. *J. Clin. Microbiol. 1:188–91*
72. Wahl, R., Cayeux, P., Derlot, E. 1962. Méningites du nouveau-né et du nourrisson à streptocoques du groupe B. *Pathol. Biol.* 10: 985–87
73. Bergqvist, G. 1974. Neonatal infections caused by group B streptococci. *Scand. J. Infect. Dis.* 6:29–31
74. Monteil, H., Rousset, A., Messer, J., Minck, R. 1975. Le streptocoque hémolytique du groupe B de Lancefield. *Nouv. Presse Med.* 4:417–18
75. Facklam, R. R., Padula, J. F., Thacker, L. G., Wortham, E. C., Sconyers, B. J. 1974. Presumptive identification of group A, B, and D streptococci. *Appl. Microbiol.* 27:107–13
76. Christie, R., Atkins, N. E., Munch-Petersen, E. 1944. A note on a lytic phenomenon shown by group B streptococci. *Aust. J. Exp. Biol. Med. Sci.* 22:197–200
77. Livoni, P., Mφlgaard, A. 1965. Gruppe B-streptokokinfektioner hos mennesker og hos kvaeg. *Ugeskr. Laeg.* 127: 1362–68
78. Lancefield, R. C. 1940–1941. Specific relationship of cell composition to biological activity of hemolytic streptococci. *Harvey Lect.* 36:251–90
79. Wilkinson, H. W., Thacker, L. G., Facklam, R. R. 1973. Nonhemolytic group B streptococci of human, bovine, and ichthyic origin. *Infect. Immun.* 7:496–98
80. Roe, M. H., Todd, J. K., Favara, B. E. 1976. Nonhemolytic group B streptococcal infections. *J. Pediatr.* 89:75–77
81. Matsen, J. M., Coghlan, C. R. 1972. See Ref 10, pp. 189–209
82. Chabbert, Y. A., Horodniceanu, T. 1975. Sensibilité des streptocoques A et B vis-à-vis de la pénicilline G, de la pénicilline V, de l'ampicilline et de l'amoxicilline. *Nouv. Presse Med.* 4: 2437–40
83. Overturf, G. D., Horowitz, M., Wilkins, J., Mathies, A. W. 1976. *Clin. Res.* 24(2):185A (Abstr.)
84. Schauf, V., Deveikis, A., Riff, L., Serota, A. 1976. Antibiotic-killing kinetics of group B streptococci. *J. Pediatr.* 89:194–98
85. Wilkinson, H. W., Facklam, R. R., Wortham, E. C. 1973. Distribution by serological type of group B streptococci isolated from a variety of clinical material over a five-year period (with special reference to neonatal sepsis and meningitis). *Infect. Immun.* 8:228–35
86. Anthony, B. F., Okada, D. M., Hobel, C. J. 1976. See Ref. 83, p. 182A
87. Aber, R. C., Allen, N., Howell, J. T., Wilkinson, H. W., Facklam, R. R. 1976. Nosocomial transmission of

group B streptococci. *Pediatrics.* 58: 346–53

88. Lancefield, R. C., Freimer, E. H. 1966. Type-specific polysaccharide antigens of group B streptococci. *J. Hyg.* 64:191–203
89. Lancefield, R. C., McCarty, M., Everly, W. N. 1975. Multiple mouse-protective antibodies directed against group B streptococci. *J. Exp. Med.* 142:165–79
90. Anthony, B. F., Concepcion, N. F., Okada, D. M., Hobel, C. J. 1974. *Proc. Intersci. Conf. Antimicrob. Agents Chemother., 14th San Francisco,* p. 253 (Abstr.)
91. Wilkinson, H. W., Eagon, R. G. 1971. Type-specific antigens of group B type Ic streptococci. *Infect. Immun.* 4:596–604
92. Anthony, B. F. 1976. Immunity to the group B streptococci: Interaction of serum and macrophages with types Ia, Ib, and Ic. *J. Exp. Med.* 143:1186–98
93. Dossett, J. H., Williams, R. C. Jr., Quie, P. G. 1969. Studies on interaction of bacteria, serum factors and polymorphonuclear leukocytes in mothers and newborns. *Pediatrics* 44:49–57
94. Klesius, P. H., Zimmerman, R. A., Mathews, J. H., Krushak, D. H. 1973. Cellular and humoral immune response to group B streptococci. *J. Pediatr.* 83:926–32
95. Baker, C. J., Kasper, D. L. 1976. Correlation of maternal antibody deficiency with susceptibility to neonatal group B streptococcal infection. *N. Engl. J. Med.* 294:753–56
96. Baker, C. J., Kasper, D. L, Paredes, A., McCormack, W. C. 1976. *Pediatr. Res.* 10(4):395 (Abstr.)
97. Hemming, V. G., Hall, R. T., Hill, H. R. 1976. *Pediatr. Res.* 10(4):399 (Abstr.)
98. Hood, M., Janney, A., Dameron, G. 1961. Beta hemolytic streptococcus group B associated with problems of the perinatal period. *Am. J. Obstet. Gynecol.* 82:809–18
99. Bergqvist, G., Hurvell, B., Thal, E., Vaclavinkova, V. 1971. Neonatal infections caused by group B streptococci. *Scand. J. Infect. Dis.* 3:209–12
100. Beargie, R., Lynd, P., Tucker, E., Duhring, J. 1975. Perinatal infection and vaginal flora. *Am. J. Obstet. Gynecol.* 122:31–33
101. Niesen, M., Müller, H. E., Schneider, J. 1973. B-streptokokken-infektionen. *Fortsch. Med.* 91:986–92
102. Sokol, R. J., Walker, R. S. 1973. β-hemolytic streptococcus in a population of antepartum patients. *Obstet. Gynecol.* 42:227–32
103. Christensen, K. K., Ripa, T., Agrup, G., Christensen, P. 1976. Group B streptococci in human urethral and cervical specimens. *Scand. J. Infect. Dis.* 8:75–78
104. Hall, R. T., Barnes, W., Krishnan, L., Harris, D. J., Rhodes, P. G., Fayez, J., Miller, G. L. 1976. Antibiotic treatment of parturient women colonized with group B streptococci. *Am. J. Obstet. Gynecol.* 124:630–34
105. Schauf, V., Hlaing, V. 1976. Group B streptococcal colonization in pregnancy. *Obstet. Gynecol.* 47:719–21
106. Baker, C. J., Barrett, F. F., Yow, M. D. 1975. The influence of advancing gestation on group B streptococcal colonization in pregnant women. *Am. J. Obstet. Gynecol.* 122:820–23
107. Wallin, J., Forsgren, A. 1975. Group B streptococci in venereal disease clinic patients. *Br. J. Vener. Dis.* 51:401–4
108. Kexel, G., Beck, K. J. 1965. Untersuchungen über die haufigkeit der B. streptokokken im wochenbett. *Geburtschilfe Frauenheilkd.* 25:1078–85
109. Christensen, K. K., Christensen, P., Flamholc, L., Ripa, T. 1974. Frequencies of streptococci of groups A, B, C, D and G in urethra and cervix swab specimens from patients with suspected gonococcal infection. *Acta Pathol. Microbiol. Scand.* 82:470–74
110. Obiger, G. 1975. Gruppe B- streptokokken im genitale der frau. *Arch. Gynäekol.* 218:65–84
111. Badri, M. S., Zawaneh, S., Cruz, A., Baer, H., Spellacy, W. N., Ayoub, E. M. 1976. *Pediatr. Res.* 10(4):394 (Abstr.)
112. Hare, R., Maxted, W. R. 1935. The classification of haemolytic streptococci from the stools of normal pregnant women and of cases of scarlet fever by means of precipitin and biochemical tests. *J. Pathol. Bacteriol.* 41:513–20
113. Smith, F. R., Sherman, J. M. 1938. The hemolytic streptococci of human feces. *J. Infect. Dis.* 62:186–89
114. Bergqvist, G., Hurvell, B., Thal, E., Vaclavinkova, V. 1976. The persistence of group B streptococci in families. *Scand. J. Infect. Dis.* 8:79–81
115. Baker, C. J., Goroff, D. K., Alpert, S., Crockett, V. A., Hayes, C., Zinner, S. H., Evard, J. R., McCormack, W. M. 1975. *Proc. Intersci. Conf. Antimicrob.*

Agents Chemother., 15th, Chicago, p. 319 (Abstr.)

116. Steere, A. C., Aber, R. C., Warford, L. R., Murphy, K. E., Feeley, J. C., Hayes, P. S., Wilkinson, H. W., Facklam, R. R. 1975. Possible nosocomial transmission of group B streptococci in a newborn nursery. *J. Pediatr.* 87:784–87
117. Winterbauer, R. H., Fortuine, R., Eickhoff, T. C. 1966. Unusual occurrence of neonatal meningitis due to group B beta-hemolytic streptococci. *Pediatrics* 38:661–62
118. Kvittingen, J. 1968. Beta-haemolytic streptococcus group B causing neonatal meningitis. *Acta Pathol. Microbiol. Scand.* 74:143–44
119. Uttley, A. H. C., Mitchison, R. S. 1975. Group B streptococci in perinatal infections. *Br. Med. J.* 3:158
120. Paredes, A., Wong, P., Yow, M. D. 1976. Failure of penicillin to eradicate the carrier state of group B streptococcus in infants. *J. Pediatr.* 89:191–93
121. Broughton, D. D., Mitchell, W. G., Grossman, M., Hadley, W. K., Cohen, M. S. 1976. Recurrence of group B streptococcal infection. *J. Pediatr.* 89:183–85
122. Truog, W. E., Davis, R. F., Ray, C. G. 1976. Recurrence of group B streptococcal infection. *J. Pediatr.* 89:185–86
123. Walker, S. H., Santos, A. Q., Quintero, B. A. 1976. Recurrence of group B III streptococcal meningitis. *J. Pediatr.* 89:187–88
124. Robbins, J. B., Schneerson, R., Argaman, M., Handzel, Z. T. 1973. *Haemophilus influenzae* type b: Disease and immunity in humans. *Ann. Intern. Med.* 78:259–69
125. Goldschneider, I., Gotschlich, E. C., Artenstein, M. S. 1969. Human immunity to the meningococcus. *J. Exp. Med.* 129:1307–26

Ann. Rev. Med. 1977. 28:371–91
Copyright © 1977 by Annual Reviews Inc. All rights reserved

PROPHYLAXIS OF INFECTIVE ENDOCARDITIS: A REEVALUATION

◆7250

James N. Sipes, M.D., Robert L. Thompson, M.D.,
and Edward W. Hook, M.D.
Department of Medicine, University of Virginia School of Medicine,
Charlottesville, Virginia 22901

INTRODUCTION

Transient bacteremia is recognized as a fundamental event in the pathogenesis of bacterial endocarditis (1). The causative organisms usually enter the circulation from the oropharynx, but other portals of entry, such as the skin and the respiratory, genitourinary, and gastrointestinal tracts, are well documented. Although infective endocarditis (IE) occurs in patients without evident preexisting heart disease, the majority of cases occur in patients with rheumatic, congenital, or other cardiovascular diseases (2, 3). The temporal association of IE with a variety of circumstances and procedures known to produce transient bacteremia, and the identification of a population at risk, have provided the rationale for the administration of antibiotics to prevent bacterial invasion of the bloodstream and multiplication of organisms at an intravascular focus.

The concept of antibiotic prophylaxis to prevent IE in patients with underlying heart diseases has been widely accepted, and guidelines and specific antibiotic regimens have been recommended by the American Heart Association (4) and others. Despite these recommendations, there has been no definitive clinical documentation of the efficacy of antimicrobials in preventing valvular infections in man. Sporadic reports of IE occurring despite prophylactic antibiotic therapy are cited as a possible cause of concern that current prophylactic regimens may be suboptimal or even ineffective (5–9). However, instances of failure of prophylaxis are few in number, with eight examples readily identified in the literature, of which only three received adequate prophylaxis by current standards.

In this paper, we reevaluate the role of chemoprophylaxis for IE. We consider the origin of bacteremias causing IE, the types of underlying cardiovascular diseases that predispose to IE after bacteremia, and the evidence that chemoprophylaxis is effective.

PORTALS OF ENTRY OF BACTERIA

Transient bacteremia is a common event. It may occur spontaneously in normal persons without evident precipitating trauma, after trauma to epithelial surfaces with an indigenous flora (e.g. dental extraction), or during the course of infection at a local site (e.g. pneumonia, prostatitis). A presumptive portal of microbial invasion in patients with IE can be demonstrated generally in about 50% of the cases (3, 10).

The oropharynx is the most common site of microbial invasion of the bloodstream when a portal of entry can be defined in patients with IE. Transient bacteremia has been a well-recognized event after dental extraction since Okell & Elliott (11) in 1935 demonstrated alpha-hemolytic streptococci in blood cultures of 34% of patients without gingivitis and in 75% of patients with severe gingival disease immediately after multiple extractions. Quantitative cultures revealed less than 15 colony-forming units (CFU) per ml in 95% of the patients. These findings were confirmed by McEntegart (12) in 1949 in a study of 200 patients, which showed bacteremia with alpha-hemolytic streptococci in more than 50% of patients after dental extractions. The frequency of bacteremia was positively correlated with severity of gingival disease, number of teeth extracted, and degree of tissue injury. In this study, quantitative cultures revealed that 66% of bacteremic patients had colony counts of less than 10 CFU per ml.

Studies of similar design utilizing modern bacteriologic techniques have confirmed and extended the early observations of Okell & Elliott (11) and McEntegart (12). For example, Crawford et al (13) in 1974 reported bacteremia with one or more organisms in 23 of 25 patients with periodontitis immediately after dental extraction. The organism most frequently isolated was *Streptococcus viridans,* but a large variety of other organisms were recovered, including coagulase-negative staphylococci; anaerobic streptococci; and *Bacteroides, Fusobacteria Veillonella* and *Vibrio sp.* In general, any microbe that inhabits the oropharynx may gain access to the circulation.

Dental manipulations other than extractions have also been shown to result in bacteremia. The incidence of bacteremia in these situations also appears directly related to the severity of preexisting periodontitis as well as the amount of tissue trauma. Cobe (14) demonstrated bacteremia in 40% of 350 patients undergoing dental cleaning, 24% of 305 persons during toothbrushing, and 17% of 225 subjects chewing hard candy, but in none of 200 persons chewing gum. The use of oral irrigation devices was shown recently to result in bacteremia in 8 of 30 patients without gingival disease (15), while another study of 30 patients with generalized periodontitis revealed that 15 had bacteremia after oral irrigation (16). Antimicrobial prophylaxis to prevent bacteremias associated with these activities of daily life is impractical. However, the relationship of bacteremia to gum disease emphasizes the importance of oral hygiene as a prophylactic measure.

The administration of penicillin immediately prior to the dental procedure decreases the incidence of positive blood cultures after the procedure (17, 18). Some of the decrease reported may have been related to in vitro action of antibiotic remaining in the blood used for culture. However, an impressive reduction in bacteremia, as indicated by positive blood cultures, is still observed if measures are taken to inhibit the action of any antimicrobial that might be transferred to the culture flask. For example, Elliott & Dunbar (19) studied retrospectively 100 children with cardiac disease who received penicillin prophylaxis prior to extraction and added penicillinase to the cultures to inactivate residual antibiotic. In this series, 36 of 100 (36%) children without cardiac disease receiving no antibiotics had streptococcal bacteremia at five minutes after extraction. In contrast, streptococcal bacteremia occurred in only 4 of 22 (18%) children who received oral phenoxymethyl penicillin, 125–250 mg every 6 hr for 3 doses prior to the procedure, and in only 2 of 51 (3.9%) receiving a single intramuscular dose of 500,000 units of benzyl penicillin prior to the procedure. The use of phenolated antiseptic gargles has been reported to decrease the incidence of postextraction bacteremia (20), but gargles should not be used without concomitant systemic antibiotic therapy.

The emergence of penicillin-resistant streptococci in the mouth and throat during prophylactic antimicrobial treatment has been well documented. Garrod & Waterworth (21) studied 74 strains of streptococci from 31 patients receiving penicillin prophylaxis for acute rheumatic fever and found that the vast majority of strains were inhibited by penicillin in a concentration of 2–32 units per ml, whereas 39 of 40 strains isolated from controls were inhibited by 0.25 units or less per ml. Other investigators have confirmed a higher prevalence of penicillin-resistant organisms in patients receiving prophylactic oral penicillin (22). These observations indicate that IE due to penicillin-resistant organisms can be anticipated more frequently in patients receiving penicillin prophylaxis (23). Chemoprophylaxis for IE in patients receiving penicillin to prevent rheumatic fever recurrences should cover organisms relatively resistant to penicillin.

The population of penicillin-sensitive streptococci decreases in the saliva one day after the initiation of penicillin therapy and is replaced by resistant organisms by the end of the second day (21). The rapid emergence of resistant organisms with the use of antibiotics indicates that prophylactic therapy for IE should precede the operative procedure only by the length of time required to attain maximal serum concentrations.

It has been suggested that edentulous patients have a lower risk of developing streptococcal endocarditis than dentulous persons (24). In fact, elective extraction of all remaining teeth after an episode of IE has been proposed (25). However, at least 17 cases of endocarditis caused by *S. viridans* in edentulous patients have been reported (26–29). The source of bacteremia in several of these patients apparently was ulcers of the gingival mucosa related to poorly fitting dentures. In patients with predisposing cardiac lesions, dental extractions should be performed for appropriate dental indications rather than as a routine procedure for prevention of IE.

Other procedures involving manipulation of the oropharynx and upper respiratory tract also result in bacteremia. Tonsillectomy has been associated with an incidence of bacteremia ranging from 20 to 38% (30, 31). Bacteremia occurs in patients undergoing nasotracheal intubation (16%) (31) and during nasotracheal suctioning (16%) (32), but apparently occurs less frequently in patients undergoing orotracheal intubation (33). Rigid-tube bronchoscopy in 52 patients resulted in bacteremia in 8 (15.4%) with organisms normally isolated from the upper respiratory tract (34). In contrast, flexible fiberoptic bronchoscopy has not been associated with bacteremia in over 150 prospectively studied procedures (35, 36). It has been suggested also that upper respiratory tract infections alone might lead to transient bacteremia, but there is no definite evidence to support this contention (1).

Upper gastrointestinal endoscopy, now performed with flexible fiberoptic instruments, is associated with significant oropharyngeal manipulation and with some trauma to the mucosa of the upper gastrointestinal tract. In one study, 4 of 50 patients (8%) had bacteremia after gastrointestinal tract endoscopy. These bacteremias were of low magnitude and due to organisms normally found in the oropharynx or small bowel, including streptococci, staphylococci, *Neisseria sp.*, and a variety of anaerobic organisms (37). Bacteremia has also been reported after endoscopic retrograde cholangiopancreatography (ERCP) (38), but there have been no prospective studies of the incidence of bacteremia with ERCP.

Procedures resulting in trauma to the colonic mucosa might also be expected to result in bacteremia with stool containing 10^{11} bacteria per g (39). Indeed, there is speculation that bacteremia may occur with minor mucosal trauma such as normal defecation (40). Four cases of enterococcal endocarditis have been reported in association with colorectal carcinoma (41, 42).

One study of patients undergoing sigmoidoscopy revealed that 19 of 200 (9.5%) had bacteremia with 15–51 CFU per ml of blood. Eleven of these 19 isolates were enterococci (43). Similarly, bacteremia was detected in 20 of 175 (11.4%) patients undergoing barium enema; nine of the 20 patients' isolates were enterococci (44). The bacteremias resulting from sigmoidoscopy and barium enema were asymptomatic and blood cultures were uniformly negative 30 min after the procedure. A recent study by Engeling and colleagues (45) has failed to confirm the results of LeFrock (43); bacteremia was not detected by Engeling and co-workers in a series of 54 patients undergoing sigmoidoscopy. In our opinion, the indications for chemoprophylaxis in patients at risk who undergo sigmoidoscopy or barium enema are not well established at the present time. If prophylaxis is given for surgical or other invasive procedures involving the lower gastrointestinal tract, it should be directed against the enterococcus, the most frequent cause of bacteremia in this setting.

Surgical procedures involving the gastrointestinal tract, particularly cholecystectomy, have been implicated as portals of entry of microorganisms of IE (3). Percutaneous liver biopsy also has been associated with bacteremia immediately after the procedure in 12 of 89 (14.4%) patients in one prospective study. Organisms included gram-negative bacilli, streptococci, enterococci, and staphylococci. In about one half of the patients, biliary tract infection was thought to be the source of bacteremia (46).

The genitourinary tract is also frequently implicated as a portal of entry of microorganisms. The anterior third of the normal urethra is colonized with bacteria, including enterococci. Thus procedures involving the urinary or genital tracts might be anticipated to result in bacteremia; indeed, the genitourinary tract is considered to be the portal of entry in over half of the patients with enterococcal endocarditis (47).

Routine urethral catheterization results in transient bacteremia in about 7% of patients (48). Removal of indwelling urethral catheters yields bacteremia in 2–22% of patients without urinary tract infections. (49, 50) and in up to 30% of patients with concurrent urinary tract infections (49). Urethral dilatation has been noted to cause bacteremia in 10–25% of patients (48, 49) and cystoscopy results in bacteremia in 19–22% of patients (48, 51). Resection of the prostate by the transurethral approach results in transient bacteremia in 33–44% of cases (48, 52). However, bacteremia after prostatectomy is much more frequent in the presence of urinary tract infection (57.7%) than in patients without infection (10.8%) (52). Retropubic prostatectomy also leads to bacteremia. In one study of 44 patients, bacteremia was detected in 68%. The majority of bacteremias occurred in patients with preexisting urinary tract infections, with 82% of those patients being bacteremic after surgery compared to only 12.8% of patients without preexisting infection. Approximately one third of the organisms isolated were enterococci (53). In another study of 82 patients, suprapubic prostatectomy resulted in bacteremia in only 7.3%; bacteremia occurred only in patients with preoperative urinary infection (54). Massage of an infected prostate can also lead to positive blood cultures (51).

In females, the vagina, the cervix, and the anterior third of the urethra are normally colonized by a variety of microorganisms, including *S. viridans*, enterococci, and staphylococci (55, 56). It is not surprising that transient bacteremia occurs after manipulation of or trauma to the tissues. Transient bacteremia occurs occasionally during normal menstrual periods (57) and during the third stage of labor (58, 59). Transient bacteremia and endocarditis have been reported during normal pregnancy (60) and after vaginal delivery (61), induced or spontaneous abortion, cesarean section (47), dilatation and curettage of the uterus (61), pelvic infection (57), and placement of intrauterine contraceptive devices (62, 63). About 20% of cases of enterococcal endocarditis result from obstetric and gynecologic procedures (47, 61, 64).

Peritoneal dialysis is associated with bacteremia in 2.5–4% of patients (65, 66). Bacteremia occurs occasionally during hemodialysis (67, 68), and endocarditis and endarteritis have been reported in hemodialysis patients (69). The increased susceptibility of hemodialysis patients to intravascular infection, even involving previously normal heart valves, appears to be related to the presence of the arteriovenous shunt (70, 71). Experimental arteriovenous fistulae in dogs are associated with a high incidence of spontaneous IE (72), and experimental IE can be produced with a smaller inoculum of organisms in the presence of an arteriovenous fistula than in dogs without fistulae (73). The elevated cardiac output resulting from arteriovenous fistulae may produce sufficient endothelial damage to evoke nonbacterial thrombus formation, which serves as a nidus for bacterial colonization (74). Because of the increased risk of intravascular infection in hemodialysis patients, with or without predisposing cardiac disease, antibiotic prophylaxis should be given to these patients just as for patients with predisposing heart lesions (75).

It was initially suggested that diagnostic cardiac catheterization might be associated with a significant incidence of bacteremia, but in a prospective study of 106 catheterizations, Sande and colleagues found no cases of bacteremia (76). Other procedures involving the cardiovascular system have been associated with bacteremia, including permanent and temporary transvenous pacemakers (77, 78), indwelling arterial catheters (79), and arterial catheter pressure transducers (80). Permanent transvenous pacemakers have also been associated with IE (78).

In the last 20 years, expansion of cardiovascular surgery has led to many instances of postoperative IE. Endocarditis after closed-heart procedures is less common than after open-heart procedures, and insertion of a prosthetic valve is followed by IE 2–4 times more frequently than other open-heart procedures (3, 10, 81). Prosthetic valve endocarditis (PVE) occurs in 2–4% of patients undergoing valve replacement (82). PVE after open-heart procedures using cardiopulmonary bypass has been attributed to several factors. First, bacteremia occurring intraoperatively is well documented, with one study (83) reporting up to 75% of cardiopulmonary bypass units harboring organisms and others (84) showing intraoperative bacteremia usually secondary to contamination of cardiopulmonary bypass equipment in 20% of patients. Second, in the postoperative period, intravascular catheters and infection of wounds, urinary and respiratory tracts, or other sites are known to result in bacteremia. The organisms isolated in these studies during and immediately after surgery are the same organisms implicated in cases of PVE occurring in the early postoperative period within eight weeks of valve replacement. These data suggest a perioperative origin of these cases of "early" PVE (85). In one interesting study of 383 patients, Ankeney (84) has shown that PVE occurred only in patients with intraoperative bacteremia; no cases of early PVE occurred in patients whose intraoperative blood cultures were negative, despite the fact that prophylactic antibiotics were not used. A third factor involved in the pathogenesis of PVE is the presence of an intracardiac foreign body, which creates a focus unusually susceptible to infection (86). Finally, surgical incision of the skin may be associated with bacteremia (87). Although staphylococci are most commonly implicated, bacteremia may be caused by any of the inhabitants of normal skin, which include *Staphylococcus aureus* and *epidermidis*, diphtheroids, and *Candida albicans.*

Definitive evidence documenting a beneficial role of antibiotics in preventing PVE is lacking. Nevertheless, the frequent occurrence of bacteremia (often staphylococcal) in the perioperative period, the high mortality (70–90%) in cases of early PVE (88–90), the recognition that antimicrobials can prevent certain infections, and the long-standing recommendations for prophylaxis in patients with heart disease have resulted in the routine use of prophylactic antibiotic therapy during insertion of heart valves. Regardless of the role of antibiotics in prophylaxis, advances in operative technique that decrease the incidence of bacteremia are potentially of equal or greater importance in preventing PVE.

Prophylactic antibiotic use during prosthetic valve insertion was first suggested because of a significant incidence of staphylococcal PVE in patients not given perioperative antibiotics (91), and many studies have reported a decrease in the overall incidence of PVE, particularly staphylococcal PVE, with the use of prophylactic antibiotics (92–94). However, none of these studies was controlled or randomized, and each utilized only historical controls. Despite claims that the incidence of staphylococcal PVE after surgery is reduced, the staphylococcus continues to account for about half of the cases, even with prophylactic use of antistaphylococcal antibiotics (88–90). Cases of PVE due to other organisms sensitive to the perioperative antibiotics also continue to occur (89, 90, 95–97).

In a controlled study, Fekety et al (96) compared placebo, penicillin, and methicillin in bypass procedures. There was no difference in the rate of postoperative infection, but PVE did not occur in any of the 213 patients.

Goodman et al (97) attempted to carry out a controlled study comparing placebo, a combination of penicillin and streptomycin, and oxacillin in preventing PVE. The placebo group was discontinued before completion of the study because of the occurrence of two cases of pneumococcal PVE. However, there was no statistically significant difference in the rate of infection or the occurrence of PVE between the three groups. There did appear to be a difference in the spectrum of causative organisms, with many organisms resistant to the antibiotics used in prophylaxis.

Conte et al (98) conducted a well-designed, controlled study comparing a single dose of intraoperative cephalothin to intraoperative plus preoperative and postoperative cephalothin in open-heart surgery. There was no difference between the groups in the incidence of minor or major infections or in mortality. However, only one case of PVE occurred in the series of 53 patients and precluded analysis of this complication.

In conclusion, available data show no clearly definable benefit of antibiotic prophylaxis in the prevention of IE after cardiovascular surgery. PVE continues to occur, even with staphylococci and other organisms that are sensitive to antibiotics used in the prophylactic regimens. The prophylactic regimens apparently do shift the spectrum of organisms toward those resistant to the antibiotics used in prophylaxis, with an increasing proportion of gram-negative bacilli and probably fungi (96–101). The mortality of PVE due to organisms other than staphylococci is equal to or greater than that of staphylococcal PVE (89, 90).

Microorganisms residing on the skin are the most common etiologic agents in the pathogenesis of endocarditis in drug addicts. Although endocarditis caused by gram-negative bacilli and fungi is several times more common in drug addicts than in the general population (8–15% and 1–4%, respectively) (102), *S. aureus* is by far the most common causative organism in addicts, isolated in 40–90% of cases (103). The source of the staphylococci was initially thought to be contaminated drugs or paraphernalia. However, studies of these articles did not reveal coagulase-positive staphylococci, but showed almost universal contamination by gram-negative organisms. Addicts with endocarditis caused by coagulase-positive staphylococci are all found to be cutaneous or nasopharyngeal carriers of the same phage type of staphylococcus as that isolated from the blood (104). It is likely, therefore, that most staphylococci causing endocarditis in drug addicts enter the bloodstream from the skin, whereas gram-negative organisms causing endocarditis in this group of patients probably gain access via contaminated drugs or related paraphernalia.

Maki et al (105) have found that intravenous cannulae, including catheters and needles, are associated with bacteremia in 2–5% of cases; bacteremia occurs in 8% of patients with plastic catheters left in place more than 48 hours. These bacteremias are caused by organisms commonly isolated from the skin, including *S. aureus,* gram-negative bacilli, and enterococci. Microbial invasion of the bloodstream is particularly common in patients receiving total parenteral nutrition. Fungi are responsible for a large percentage of these cases. Intravenous cannulae have been responsible for IE, as well as for disseminated fungal infections and fatal septicemia.

Intravenous fluids, as well as 1–6% of blood products, have been reported to harbor bacteria (106), and as many as 20% of pooled platelet transfusions may be contaminated (107). Fatal septicemia has occurred with these infusions, but IE has not been reported, perhaps because gram-negative bacilli are usually isolated and they are less likely to produce IE than are streptococci or staphylococci. Intravenous therapy should be avoided whenever possible in the patient at risk for IE and, when necessary, infection-control measures such as those outlined by Goodman et al (106) should be followed to reduce the risk of bacteremia.

When considering potential sources of bacteremia in persons with endocarditis, it should be recalled that spontaneous bacteremia with small numbers of organisms occurs without obvious trauma in apparently healthy persons (108, 109). Bacteremia also occurs spontaneously in patients with minor infections such as periodontitis, periapical abscesses, and tonsillitis (11, 110, 111). Invasion of the bloodstream also occurs frequently in the course of other bacterial infections; for example, approximately 30% of patients with pneumococcal pneumonia are found to be bacteremic (112).

UNDERLYING CARDIOVASCULAR DISEASE

Rheumatic heart disease remains the most common form of underlying cardiovascular disease in patients developing IE, accounting for 40–60% of all cases, despite a remarkable decline in incidence over the past 30–40 years (2, 3). Congenital heart disease is found in 7–16% of patients with IE in recent series (113, 114). Small ventricular septal defect, patent ductus arteriosus, tetralogy of Fallot, pulmonic stenosis, coarctation of the aorta, and bicuspid aortic valve appear as predisposing lesions relatively often, whereas ostium secundum atrial septal defect, large ventricular septal defect, and other congenital abnormalities with low pressure differentials are rarely encountered with IE (115). Whereas the incidence of rheumatic heart disease has declined as an underlying condition in IE, the occurrence of so-called degenerative heart diseases, including calcific and atherosclerotic changes in the aortic and mitral valves, has apparently increased. One review reported that 25 of 100 patients with IE had degenerative heart lesions as predisposing abnormalities; if only patients over the age of 60 were considered, 50% had predisposing degenerative disease (116). This reflects the fact that nearly all severe valvular aortic stenosis in patients over the age of 65 is due to degenerative changes in the aortic valve (117, 118). Similarly, calcification of the mitral annulus predisposes to IE (119). IE also occasionally complicates myocardial infarction, involving the mural thrombus overlying the infarction (120), or developing at the site of a ventricular septal perforation after myocardial infarction (121).

Other cardiovascular conditions that have been complicated by IE include idiopathic hypertrophic subaortic stenosis (122, 123), mitral valve prolapse click-murmur syndrome (124), Marfan's syndrome (125), and syphilitic heart disease (113). Prosthetic intracardiac devices (i.e. valves, patches, and pacemakers) can be the site of IE, as previously discussed. Also, as previously noted, arteriovenous fistulae or shunts predispose to IE, irrespective of whether the patient has underlying heart disease. However, there is evidence to suggest that prophylaxis is not required after the successful repair of atrial septal defects (125), patent ductus arteriosus (4), and probably arteriovenous fistulae of all types.

In patients who have had one episode of IE, the incidence of a second episode is impressively high, perhaps in the range of 8–9% with long-term follow-up. One fourth of the patients with a second attack will have third episodes of IE (126, 127). This high risk of recurrence makes a previous episode of IE a strong indication for antibiotic prophylaxis, even without a recognizable underlying cardiovascular defect.

When all etiologies are considered, some 20–40% of patients with IE have no apparent underlying heart disease. Some of these undoubtedly have unrecognized or undiagnosable cardiac diseases, although IE definitely can occur on previously normal valves, especially after bacteremia with organisms of high virulence.

RISK OF ENDOCARDITIS AFTER BACTEREMIA

Several attempts have been made to estimate the risk of IE in susceptible persons after dental extraction, an event known to be regularly associated with bacteremia. However, assessment of risk of IE after bacteremia is difficult. Extrapolations based on numerous assumptions, including estimating the incidence of IE in the susceptible population, the number of dental extractions that patients at risk undergo, and the number of patients in which dental extraction is the precipitating event of IE, have led to the guess that the incidence of IE after dental extraction may range from 1 case per 500–3000 extractions (1). The risk of endocarditis is influenced by many variables, including type of underlying cardiovascular lesion, number and type of organisms gaining access to the blood, and immunologic status of the host. It is apparent, considering the frequency of bacteremia, that intravascular localization of blood-borne microbes is a rare event.

Garrison & Freedman (128) developed a model of IE in rabbits that has proven reproducible and practical and has allowed extensive study of the pathogenesis of IE. Trauma to the endocardial surfaces is produced by passing a polyethylene catheter transvenously or transarterially across the tricuspid or aortic valve. Nonbacterial thrombotic endocarditis is consistently produced at the site of contact between the catheter and the endocardium. The thrombus apparently serves as a site for the localization and growth of bacteria that lead to the formation of a typical vegetation. Studies using the rabbit model have demonstrated that development of IE after bacteremia is dependent on a number of factors. First, the incidence of endocarditis appears to be directly related to the amount of valvular trauma. With lesser degrees of trauma or more time for healing of induced trauma, the risk of IE is reduced (129). Trauma need not be direct, since the creation of an arteriovenous fistula, resulting in only minor endothelial damage, presumably secondary to increased cardiac output, reduces the number of organisms required to produce IE (31, 129). Additionally, left-sided experimental endocarditis is more readily established and maintained than right-sided valvular infections, perhaps because of the higher left-heart pressures. Second, the number of organisms inoculated is an important determinant of the occurrence of IE. Third, the duration of bacteremia is important, and any maneuver that decreases the clearance of bacteria from the circulation (e.g. reticuloendothelial blockade or the creation of an arteriovenous fistula) results in an increased incidence of IE (73, 74). Fourth, the microorganism involved is of major importance. In the rabbit model, the inoculum required to produce IE in 50% of animals varies from $10^{3.75}$ CFU for *S. aureus* to $10^{4.67}$ CFU for *S. viridans*, and to as high as $10^{6.29}$ for *Escherichia coli* (130).

These experimental data appear to agree with clinical impressions regarding factors influencing the risk of IE in man. As in the experimental model, predisposing cardiovascular lesions that seem to result in more endothelial trauma carry a higher risk of IE (e.g. ventricular septal defect with high pressure differential has a higher incidence of IE than low–pressure gradient atrial septal defects) (114, 115). The incidence of left-sided IE is also higher than that of right-sided endocarditis, although this may reflect the more frequent occurrence of rheumatic heart disease involving the left side of the heart, as well as other fundamental differences in the susceptibility of the chambers to bacterial infection.

The number of organisms recovered from the blood immediately after procedures causing transient bacteremia is generally relatively small, often less than 10–20 organisms per ml of blood. Organisms are cleared rapidly and bacteremia is resolved by 30 min (57). The development of endocarditis is dependent on the localization of a viable bacillus with growth potential in the small fibrin clot formed at the site of trauma. The low level of transient bacteremia and its rapid resolution do not favor intravascular localization in most instances.

The importance of the type of organism producing bacteremia is apparent in human beings and in experimental animals. The propensity for certain organisms to cause IE is related to the frequency of bacteremia with the organism and, of even greater importance, it is related to specific characteristics of the different organisms, including adhesiveness to endothelium and susceptibility to killing by antibody and complement.

Despite the multitude of organisms that can produce bacteremia, streptococci and staphylococci account for 90–95% of all IE, with the exception of PVE and endocarditis in drug addicts. Although organisms such as anaerobes or gram-negative enteric bacilli frequently cause bacteremia, they are unusual causes of IE (3). Many of the latter organisms are killed by antibody and complement, which may account for the much larger inoculum required to produce endocarditis with these organisms. Also, bacteremia of sufficient magnitude to produce endocarditis might result in endotoxic shock and death prior to the establishment of IE. The risk of IE after staphylococcal bacteremia is greater than after bacteremia with most other organisms, as witnessed by the fact that IE occurs in a significant proportion of patients with staphylococcal bacteremia irrespective of the presence of underlying heart disease (131).

In summary, many factors that alter the risk of IE after experimental bacteremia are recognized. Although many parallels exist between the naturally occurring disease in man and the experimental disease in animals, calculation of risk of IE after bacteremia in man is almost impossible. Asymptomatic transient bacteremia in man is exceedingly common, whereas endocarditis in susceptible persons is unusual. The risk of IE after bacteremia in susceptible persons is obviously small.

PREVENTION OF EXPERIMENTAL ENDOCARDITIS

Durack and co-workers (132, 133) have utilized the rabbit model to test the efficacy of various prophylactic antibiotic regimens against IE caused by streptococci. In these experiments, a catheter passed across the tricuspid or aortic valve produces sterile vegetations that serve as a nidus for microbial localization and growth after intravenous injection of 10^8 CFU of a strain of *S. viridans.* Bacterial colonization of the thrombus was uniformly observed when animals were sacrificed 24 hr after injection. Thirty minutes prior to injection of streptococci, various antibiotics were administered parenterally to evaluate their ability to prevent localization and growth of bacilli in the thrombus.

Although the streptococcus used in these experiments was highly sensitive to penicillin in vitro, with a minimum inhibitory concentration of 0.01 μg/ml, it was difficult to prevent colonization of the thrombus with a single dose of penicillin G. Aqueous penicillin G given 30 min prior to bacteremia, in single doses as large as 150 mg/kg, was not effective in 19 of 21 animals. However, when doses of this magnitude were given every 4 hr for 6 doses, 3 of 4 animals had sterile vegetations when sacrificed. A combination of penicillin G (150 mg/kg) with benzathine penicillin (7.5 mg/kg) or procaine penicillin ($\geq$50 mg/kg), which gave more prolonged serum levels, was uniformly successful in preventing valvular infection. These results were interpreted as suggesting the need for a high initial serum level as well as relatively prolonged duration of serum penicillin activity.

Vancomycin in a single dose (30 mg/kg) was uniformly successful in preventing colonization of the sterile vegetations, despite disappearance of the drug from the serum within 3 hr. Similarly, a single dose of the synergistic combination of penicillin G (150 mg/kg) and streptomycin (15 mg/kg) was uniformly effective. With vancomycin alone or with the combination of penicillin and streptomycin, quite rapid rates of killing in vitro were observed, which may explain their effectiveness in vivo. Cephazolin (30 mg/kg) every 6 hr plus streptomycin, both for 2 days, was also effective. Phenoxymethyl penicillin (30 mg/kg initially, followed by 7.5 mg/kg every 6 hr for 48 hr) was the only effective regimen that could potentially be given orally to man. All of the bacteriostatic agents tested, including tetracycline, erythromycin, and clindamycin, were ineffective in the experimental model, even when given in multiple doses.

Direct extrapolation of data obtained from the rabbit to the development of prophylactic regimens for use in humans has been criticized from several standpoints. In the experimental model, 10^8 CFU are injected under conditions that lead to intravascular localization of viable bacilli in almost all animals. The number of organisms in the experimental infection in rabbits is considerably greater than that generally involved in human transient bacteremia. Furthermore, IE develops only rarely after transient bacteremia in humans with vascular lesions. Consequently, regimens based on the animal model, while admittedly providing a wide margin of safety, probably overestimate the dose of antimicrobials required to prevent IE in humans with heart disease undergoing procedures associated with bacteremia.

Experimental data regarding the chemoprophylaxis of experimental IE caused by enterococci, staphylococci, or other organisms are not available.

PREVENTION OF IE IN MAN

The ideal approach to the prevention of IE would be prevention or correction of the underlying cardiovascular defects. Failing this, other modes of prophylaxis must be sought. Patients at risk should avoid procedures known to be associated with bacteremia whenever possible. Certain nonessential procedures, such as intrauterine devices for contraception, should be avoided. Oral hygiene should be maintained at the highest possible level. Special effort should be made to insure that necessary invasive procedures involving contaminated epithelial surfaces or local sites of infection are carried out with minimal trauma. For example, it is known that endodontic procedures are less likely to be associated with bacteremia than tooth extraction or gingival surgery (134). Existing infection should be eliminated prior to manipulation—for example, the urinary tract infection prior to genitourinary procedures—to decrease the probability of bacteremia. Intravascular and genitourinary catheters should be avoided, but if their use is mandatory, they should be removed as soon as possible. Other infections (e.g. pneumonia, cutaneous infections) should be treated promptly with antibiotics to decrease the likelihood of bacteremia.

Chemoprophylaxis with antibiotic agents is effective in preventing experimental endocarditis in rabbits and may be effective in humans by decreasing the frequency and magnitude of bacteremia associated with traumatic procedures, by eliminating bacteria that have gained access to the bloodstream, or by eradicating bacteria lodged on heart valves before endocarditis is established. Despite the fact that definitive evidence of efficacy of antibiotics in the prevention of endocarditis in man is lacking, there appear to be sufficient data to warrant continued use of chemoprophylaxis. In addition to data documenting efficacy in experimental IE, reported instances of IE in persons at risk who are treated with antimicrobials, as recommended by the American Heart Association, are remarkably few in number. Furthermore, current ethical attitudes and the exceedingly low incidence of IE in susceptible individuals after procedures known to evoke transient bacteremia make prospective comparative clinical study of placebo versus antibiotic difficult.

Chemoprophylaxis should be given to all persons with recognized predisposing cardiovascular conditions prior to procedures known to be associated with endocarditis and to produce bacteremia frequently. However, even if one assumes that antimicrobial prophylaxis prevents essentially all cases of IE, a substantial number of cases would continue to occur. As many as 40% of patients with IE have no previously recognized underlying cardiovascular disease and therefore would not receive prophylaxis. Furthermore, even in patients known to be at risk, more than half of bacteremias leading to IE originate from unrecognized sources and would not be prevented.

Patients with underlying cardiovascular disease should be informed of their predisposition to infection and of the importance of chemoprophylaxis. Several studies have confirmed that 70–80% of patients at risk and their families are unaware of the potential danger of IE after invasive procedures (22, 23). It is also important that health professionals be diligent in identifying the patients at risk and in delivering prophylactic regimens. In a study of 2069 patients attending a dental clinic, 113 patients were identified as at risk of IE because of cardiac diseases. Sixty-four percent of these patients indicated that their dentists did not inquire about a history of cardiac disease, and 77% of 424 dental procedures were performed without antibiotic prophylaxis (17).

There is general agreement that chemoprophylaxis should be initiated just minutes prior to the procedure so as to achieve an active serum level of antimicrobials at the time of bacteremia and to minimize the risk of the selection of a resistant flora by prolonged therapy prior to the procedure. Prophylaxis should be directed against the organism most likely to be responsible for bacteremia in the specific clinical setting.

Patients with predisposing cardiovascular disease who undergo traumatic procedures involving the oropharynx, including dental extraction, dental prophylaxis, other forms of oral surgery, and rigid-tube bronchoscopy, should receive antibiotic prophylaxis primarily directed against streptococci of the viridans type. Several regimens recommended by the American Heart Association (4) for procedures involving the oropharynx have been used extensively:

1. Procaine penicillin G, 600,000 units, mixed with crystalline penicillin G, 200,000 units, intramuscularly (i.m.) 1 hr prior to the procedure and once daily for 2 days after the procedure.
2. Phenoxymethyl penicillin (penicillin V) or phenethicillin, 500 mg orally (p.o.) 1 hr prior to the procedure and then 250 mg p.o. every 6 hr for the remainder of that day and for 2 days after the procedure.
3. Penicillin G, 1,200,000 units p.o., 1 hr prior to the procedure and then 600,000 units p.o. every 6 hr for the remainder of that day and for 2 days after the procedure.
4. In patients allergic to penicillin or who have been receiving penicillin prophylaxis against rheumatic fever, erythromycin stearate, 500 mg p.o. in adults (20 mg/kg in children) 1½–2 hr before the procedure and then 250 mg p.o. in adults (10 mg/kg in children) every 6 hr for the remainder of that day and for 2 days after the procedure.

During recent years, it has been possible to test these regimens in experimental IE in rabbits. The antimicrobials in dosages recommended above were ineffective under these experimental conditions in "preventing endocarditis;" that is, 24 hr after initiation of bacteremia with 10^8 CFU in rabbits, viable streptococci could still be isolated from intravascular thrombi. However, with larger doses of penicillin, with combination of penicillin and streptomycin, or with vancomycin, it was possible to achieve more consistent sterilization of the vegetations. On the basis of these observations in rabbits, several groups have recommended that the prophylactic regimens for man be revised. Durack & Petersdorf (132, 133) recommend the following regimens:

1. Aqueous penicillin G, 2,000,000 units, procaine penicillin G, 600,000 units, and streptomycin, 1.0 g, by i.m. injection 30 min prior to the procedure.
2. Vancomycin, 0.5 g, intravenously (i.v.) 30 min prior to the procedure.
3. Cephazolin, 1.0 g, plus streptomycin, 0.5 g, i.m. 30 min before the procedure and every 8 hr for two additional doses;
4. Phenoxymethyl penicillin, 2.0 g, p.o. 30 min before the procedure followed by 0.5 g p.o. every 6 hr for 48 hr.

Kaye (87) recommends the addition to the first regimen of 600,000 units of procaine penicillin i.m. on each of the two days after the procedure. He also suggests that the vancomycin and cephazolin-streptomycin regimens be reserved for patients hypersensitive to penicillin and that these regimens be continued for a total of 72 hr. All of these regimens are bactericidal and provide sufficiently high initial serum concentrations and duration of activity required to prevent uniform colonization of the vegetation in rabbits. In patients receiving penicillin for rheumatic fever prophylaxis, the chemoprophylactic regimen chosen should be bactericidal for streptococci relatively resistant to penicillin and might consist of any of the first three regimens mentioned above. If oral prophylaxis is required in a penicillin-allergic patient, Kaye recommends erythromycin, 500 mg, 4 times daily for 72 hr, but such nonbactericidal agents are less effective in the rabbit model.

If one assumes that antimicrobial prophylaxis is effective in preventing IE in man, then one can also assume that the larger doses of bactericidal antimicrobials recommended by Durack & Petersdorf (132, 133) and Kaye (87) will offer a margin of safety that exceeds that assumed to exist with the regimens previously recommended by the American Heart Association (4). However, there is a question as to whether the revised recommendations are necessary. As pointed out earlier, there are very few instances of documented failure of current regimens to prevent IE in humans at risk. Furthermore, the comparability of the animal model to IE in man is quantitatively remote; that is, in the experimental model the variables are controlled so that virtually every rabbit develops IE after injection of bacteria, whereas in man, IE rarely follows bacteremia. It is possible that, if the conditions in man were reproduced in rabbits, the regimens recommended by the American Heart Association would be found effective in preventing IE in man.

Prophylactic antibiotic therapy for genitourinary and lower gastrointestinal tract procedures should be primarily directed against the enterococcus. In this instance, there are no experimental data to suggest dosage or duration of therapy. The regimens recommended by the American Heart Association (4) for prophylaxis for gastrointestinal and genitourinary tract surgery and instrumentation are as follows:

1. Procaine penicillin G, 600,000 units, mixed with crystalline penicillin G, 200,000 units, plus streptomycin, 1–2 g, injected i.m. 1 hr prior to and once daily for 2 days after the procedure (for children, the dose of streptomycin is 40 mg/kg, not to exceed 1 g per 24 hr).
2. Ampicillin, 25–50 mg per kg, p.o. or i.v. 1 hr before the procedure and then 25 mg/kg every 6 hr for the remainder of that day and for 2 days after the procedure, plus streptomycin as above.
3. In patients allergic to penicillin, erythromycin can be given p.o. as described above for prevention of *S. viridans* infections, plus streptomycin as above; an alternative regimen for the penicillin-allergic patient is vancomycin, 0.5–1.0 g i.v. (20 mg/kg in children) 1 hr prior to the procedure and then 0.5 g i.v. (10 mg/kg in children) every 6 hr for 72 hr, plus streptomycin as above.

Extrapolating from the animal work with *S. viridans,* Kaye (87) feels that it is unlikely that these regimens of penicillin, oral ampicillin, or erythromycin would be effective in preventing enterococcal endocarditis. He proposes that therapeutic doses of antibiotics for established enterococcal endocarditis would be effective and provide a wide margin of safety. Therefore, the following three regimens have been recommended by Kaye (87):

1. Aqueous penicillin G, 20,000,000 units given i.v. over 24 hr plus streptomycin, 1.0 g, i.m. every 12 hr beginning 1 hr before the procedure and continuing for a total of 72 hr.
2. Aqueous penicillin G as above, plus gentamicin, 1.7 mg/kg, i.m. every 8 hr beginning 1 hr before the procedure for a total of 72 hr.
3. Vancomycin, 0.5 g, given i.v. every 6 hr plus streptomycin or gentamicin as above beginning 1 hr before the procedure for a total of 72 hr.

Although these regimens for prevention of IE after genitourinary procedures may offer wide margins of safety, they are not likely to meet with wide acceptance. First, the evidence indicates that very few cases of enterococcal endocarditis have been reported in patients at risk who have received one of the regimens recommended by the American Heart Association (4). Second, the regimens recommended by Kaye would require hospitalization for three days in most instances. A third factor to be considered is the increased cost resulting from large doses of parenteral antimicrobials and the hospitalization. In our opinion, better definition of risks and benefits than exists at present should be obtained prior to discarding the recommendations of the American Heart Association.

The surgical manipulation of normal or infected cutaneous tissues requires chemoprophylaxis directed against *S. aureus.* A penicillinase-resistant penicillin, such as nafcillin, or cephalothin in dosage of 2 g given i.v. every 4 hr, or cephazolin, 1 g, i.m. every 8 hr, beginning 1 hr before the procedure and continuing for 72 hr, may be used. Parenteral antimicrobials are advisable when the probability of staphylococcal bacteremia is high; however, if oral prophylaxis is required, dicloxacillin or cephalexin, 500 g every 6 hr, can be used for 72 hr after the procedure.

Despite the lack of conclusive evidence of efficacy, the use of prophylactic antibiotics with cardiovascular surgery is almost universal and undoubtedly will continue. Prophylaxis should be directed primarily against the staphylococcus, as it is the most common cause of PVE. Parenteral antistaphylococcal agents, such as nafcillin, cephalothin, or vancomycin, should be given in full therapeutic doses i.v. beginning 1 hr prior to surgery and continuing intraoperatively and postoperatively for 3 days or until all intravascular lines and catheters are removed.

SUMMARY

Established practice dictates that physicians and dentists give antimicrobial agents at the time of procedures associated with bacteremia in an effort to prevent IE in persons with underlying cardiovascular diseases. Although prospective controlled study has not established efficacy, very few cases of IE have been reported in patients at risk who have been treated with regimens recommended by the American Heart Association. Furthermore, studies in rabbits in recent years have shown that IE can be prevented by antimicrobials. The results in rabbits have raised some theoretical questions regarding efficacy of currently recommended regimens for man and have led some investigators to propose revisions of the American Heart Association recommendations. We feel that direct translation of results from experimental IE in rabbits into recommendations for prophylaxis of IE in man is subject to question.

It is unlikely that a prospective study comparing placebo with antimicrobials will be undertaken, but it is possible and reasonable to undertake studies to document the occurrence of IE in patients at risk who are given established antibiotic regimens. With such data, modifications of established regimens could be proposed and tested. Regimens that are found to be totally effective in a large number of patients at risk might be modified toward simpler programs—for example, modifying parenteral therapy to oral therapy or changing multiple doses of antibiotic to a single dose. If regimens fail, even once in several thousand patients at risk, the factors responsible might be determined and a clear rationale for modification achieved.

Literature Cited

1. Hook, E. W., Kaye, D. 1962. Prophylaxis of endocarditis. *J. Chronic Dis.* 15:635–46
2. Weinstein, L., Rubin, R. H. 1973. Infective endocarditis—1973. *Prog. Cardiovasc. Dis.* 16:239–74
3. Lerner, P. I., Weinstein, L. 1966. Infective endocarditis in the antibiotic era. *N. Engl. J. Med.* 274:199–206
4. American Heart Association Committee on Rheumatic Fever and the Committee on Congenital Heart Defects. 1972. Prevention of bacterial endocarditis. *Circulation* 46:Suppl.V, pp. 3–6
5. Thill, C. J., Meyer, O. O. 1947. Experiences with penicillin and dicumarol in the treatment of subacute bacterial endocarditis. *Am. J. Med. Sci.* 213:300–7
6. Cates, J. E., Christie, R. V., Garrod, L. P. 1951. Penicillin-resistant subacute bacterial endocarditis treated by a combination of penicillin and streptomycin. *Br. Med. J.* 1:653–55
7. Dormer, A. E. 1958. Bacterial endocarditis: survey of patients treated between 1945 and 1956. *Br. Med. J.* 1:63–69
8. Barnes, C. G., Hurley, R. 1963. Antibiotic cover for dental extractions. *Br. Med. J.* 2:1205
9. Durack, D. T., Littler, W. A. 1974. Failure of "adequate" penicillin therapy to prevent bacterial endocarditis after tooth extractions. *Lancet* 2:846–47

10. Hilson, G. R. F. 1970. Is chemoprophylaxis necessary? *Proc. R. Soc. Med.* 63:267–71
11. Okell, C. C., Elliott, S. D. 1935. Bacteremia and oral sepsis with special reference to the etiology of subacute endocarditis. *Lancet* 2:869–72
12. McEntegart, M. G., Porterfield, J. S. 1949. Bacteremia following dental extractions. *Lancet* 2:596–98
13. Crawford, J. J., Sconyers, J. R., Moriarty, J. D., King, R. C., West, J. F. 1974. Bacteremia after tooth extraction studied with the aid of prereduced anaerobically sterilized culture media. *Appl. Microbiol.* 27:927–32
14. Cobe, H. M. 1954. Transitory bacteremia. *Oral Surg.* 7:609–15
15. Berger, S. A., Weitzman, S., Edberg, S. C., Casey, J. I. 1974. Bacteremia after the use of an oral irrigation device. *Ann. Intern. Med.* 80:510–11
16. Felix, J. E., Rosen, S., App, G. R. 1971. Detection of bacteremia after the use of an oral irrigation device in subjects with periodontitis. *J. Periodontol.* 42:785–87
17. Glazer, R. J., Dankner, A., Mathes, S. B., Harford, C. F. 1948. Effect of penicillin on the bacteremia following dental extraction. *Am. J. Med.* 4:55–65
18. Hirsh, H. L., Vivino, J. J., Merril, A., Powling, H. F. 1941. Effect of prophylactically administered penicillin on incidence of bacteremia following extraction of teeth. *Arch. Intern. Med.* 81:868–78
19. Elliott, R.H., Dunbar, J. M. 1968. Streptococcal bacteremia in children following dental extractions. *Arch. Dis. Child.* 43:451–54
20. Jones, J. C., Cutcher, J. L., Goldberg, J. R., Lilly, G. E. 1970. Control of bacteremia associated with extraction of teeth. *Oral Surg.* 30:454–59
21. Garrod, L. P., Waterworth, P. M. 1962. The risks of dental extraction during penicillin therapy. *Br. Heart J.* 24: 39–46
22. McGowan, D. A., Tuohy, O. 1968. Dental treatment of patients with valvular heart disease. *Br. Dent. J.* 124: 519–20
23. Caldwell, R. L., Hurwitz, R. A., Girod, D. A. 1971. Subacute bacterial endocarditis in children. *Am. J. Dis. Child.* 122:312–15
24. Hobson, F. G., Juel-Jensen, B. E. 1956. Teeth, *Streptococcus viridans* and subacute bacterial endocarditis. *Br. Med. J. 2:1501–5*
25. Cooke, B. E. D. 1970. Relation between oral disease and acute bacterial endocarditis. *Proc. R. Soc. Med.* 63: 263–66
26. Croxson, M. S., Altmann, M. M., O'Brien, K. P. 1971. Dental status and recurrence of *Streptococcus viridans* endocarditis. *Lancet* 1:1205–7
27. Simon, D. S., Goodwin, J. F. 1971. Should good teeth be extracted to prevent *Streptococcus viridans* endocarditis? *Lancet* 1:1207–9
28. Goodman, J. S., Kalhouse, J. F., Koenig, M. G. 1973. Recurrent endocarditis due to *Streptococcus viridans* in an "edentulous" man. *South. Med. J.* 66:352–58
29. Massaro, D., Katz, S. 1960. Subacute bacterial endocarditis due to *Streptococcus viridans* in an edentulous man. *N. Engl. J. Med.* 263:911
30. Rhoades, P. S., Sibley, J. R., Billings, C. E. 1955. Bacteremia following tonsillectomy. *J. Am. Med. Assoc.* 157:877–81
31. Berry, F. A., Blankenbaker, W. L., Ball, C. G. 1973. A comparison of bacteremia occurring with nasotracheal and orotracheal intubation. *Anesth. Analg. Cleveland* 52:873–76
32. LeFrock, J. L., Klainer, A. S., Wu, W., Turndorr, H. 1974. Transient bacteremia associated with nasotracheal suctioning. *Clin. Res.* 22:646a (Abstr.)
33. Berry, F. A., Yarbrough, S., Yarbrough, N., Russell, C. A., Carpenter, M. A., Hendley, J. O. 1973. Transient bacteremia during dental manipulation in children. *Pediatrics* 51:476–79
34. Burman, S. O. 1960. Bronchoscopy and bacteremia. *J. Thorac. Cardiovasc. Surg.* 40:635–39
35. Kane, R. C., Cohen, M. H., Fossiecki, B. E. Jr., Tvardzik, A. V. 1975. Absence of bacteremia after fiberoptic bronchoscopy. *Am. Rev. Respir. Dis.* 111:102–4
36. Pereira, W., Kovnat, D. M., Khan, M. A., Iacovino, J. R., Spivack, M. L., Snider, G. L. 1975. Fever and pneumonia after flexible fiberoptic bronchoscopy. *Am. Rev. Respir. Dis.* 112:59–64
37. Shull, H. J., Greene, B. M., Allen, S. D., Dunn, G. D., Schenker, S. 1975. Bacteremia with upper gastrointestinal tract endoscopy. *Ann. Intern. Med.* 83:212–14
38. Bilbao, M. K., Dotter, C. T., Lee, T. G., Katon, R. M. 1976. Complications of endoscopic retrograde cholangiopancreatography (ERCP): a study of 10,000 cases. *Gastroenterology* 70: 314–20

39. Levison, M. E., Kaye, D. 1969. Fecal flora in man: effect of cathartic. *J. Infect. Dis.* 119:591–96
40. Weinstein, L. 1975. Chemoprophylaxis for bacterial endocarditis (Letter). *N. Engl. J. Med.* 292:427–28
41. McCoy, W. C., Mason, J. M. 1951. Enterococcal endocarditis associated with carcinoma of the sigmoid: report of a case. *J. Med. Assoc. State Ala.* 21: 162–66
42. Roses, D. F., Richman, H., Localio, S. A. 1974. Bacterial endocarditis associated with colorectal carcinoma. *Ann. Surg.* 179:190–91
43. LeFrock, J. L., Ellis, C. A., Turchik, J. B., Weinstein, L. 1973. Transient bacteremia associated with sigmoidoscopy. *N. Engl. J. Med.* 289:467–69
44. LeFrock, J. L., Ellis, C. A., Klainer, A. S., Weinstein, L. 1975. Transient bacteremia associated with barium enema. *Arch. Intern. Med.* 135:835–37
45. Engeling, E. R., Eng, B. F., Sullivan-Sigler, N., Bartlett, J. G., Gorbach, S. L. 1976. Bacteremia after sigmoidoscopy: another view. *Ann. Intern. Med.* 85:77–78
46. LeFrock, J. L., Ellis, C. A., Turchik, J. B., Weinstein, L. 1973. Transient bacteremia associated with percutaneous liver biopsy. *Clin. Res.* 21:843
47. Mandell, G. L., Kaye, D., Levison, M. E., Hook, E. W. 1970. Enterococcal endocarditis: an analysis of 38 patients observed at the New York Hospital-Cornell Medical Center. *Arch. Intern. Med.* 125:258–64
48. Sullivan, N. M., Sutter, V. L., Carter, W. T., Attebery, H. R., Finegold, S. M. 1971. Bacteremia following genitourinary tract manipulation. *Bacteriol. Proc.* M201:98 (Abstr.)
49. Slade, N. 1958. Bacteremia and septicemia after urological operations. *Proc. R. Soc. Med.* 51:331–34
50. Plorde, J. J., Kennedy, R. D., Bourne, H. H. 1965. Course and prognosis of prostatectomy: with a note on the incidence of bacteremia and effectiveness of chemoprophylaxis. *N. Engl. J. Med.* 272:269–77
51. Last, P. M., Harbison, P. F., Marsh, J. A. 1966. Bacteremia after urological instrumentation. *Lancet* 1:74–76
52. Creevy, L. D., Feeney, M. J. 1954. Routine use of antibiotics in transurethral prostatic resection. *J. Urol.* 71:615–23
53. Marshall, A. 1961. Bacteremia following retropubic prostatectomy. *Br. J. Urol.* 33:25–33
54. Steyn, J. H., Logie, N. J. 1962. Bacteremia following prostatectomy. *Br. J. Urol.* 34:459–62
55. Hunter, L. A., Long, K. R. 1958. A study of the microbiological flora of the vagina. *Am. J. Obstet. Gynecol.* 75: 865–71
56. Guilbeau, J. A. Jr., Schaub, I. G., Andrews, M. C. 1949. Penicillin therapy in the obstetrical patient: a study of its effect on the bacterial flora of the postpartum uterus. *Am. J. Obstet. Gynecol.* 58:101–9
57. Bennett, I. L. Jr., Beeson, P. B. 1954. Bacteremia: a consideration of some experimental and clinical aspects. *Yale J. Biol. Med.* 26:241–62
58. Redleaf, P. D., Fadell, E. J. 1959. Bacteremia during parturition. *J. Am. Med. Assoc.* 169:1284-85
59. Burwell, L. S., Metcalfe, J. 1958. *Heart Disease and Pregnancy: Physiology and Management.* Boston: Little, Brown
60. Crockett, E. J., Batchelor, T. M., Corbeil, J. 1960. Pregnancy and subacute bacterial endocarditis: report of a case. *Obstet. Gynecol.* 16:93–95
61. Lein, J. N., Stander, R. W. 1959. Subacute bacterial endocarditis following obstetric and gynecologic procedures. *Obstet. Gynecol.* 13:568–73
62. Cobbs, C. G. 1973. IUD and endocarditis. *Ann. Intern. Med.* 78:451
63. DeSweit, M., Ramsay, I. D., Rees, G. M. 1975. Bacterial endocarditis after insertion of intrauterine contraceptive device. *Br. Med. J.* 3:76–77
64. Geraci, J. E., Martin, W. J. 1954. Antibiotic therapy of bacterial endocarditis VI: subacute enterococcal endocarditis: pathologic and therapeutic consideration of 33 cases. *Circulation* 10:173–94
65. Stevens, R. E., Baskin, S., Greene, J. A., Weller, J. M. 1964. Peritoneal dialysis in the management of chronic renal failure. *J. Am. Med. Assoc.* 190:102–4
66. Montgomerie, J., Kalmanson, G. M., Guze, L. B. 1968. Renal failure and infection. *Medicine* 47:1–32
67. Report to the Public Health Laboratories by the working party on hemodialysis units. 1968. Infection risks of hemodialysis: some preventive aspects. *Br. Med. J.* 3:454–60
68. Robinson, P. J. A., Rosen, S. M. 1971. Pyrexial reactions during hemodialysis. *Br. Med. J.* 1:528–30
69. Hook, E. W., Wainer, H. S., McGee, T. J. 1957. Acquired arteriovenous fistula with bacterial endarteritis and endocarditis. *J. Am. Med. Assoc.* 162:1450–54

70. King, L. H., Bradley, K. P., Shires, D. L., Donohue, J. P., Glover, J. L. 1971. Bacterial endocarditis in chronic hemodialysis patients: a complication more common than previously suspected. *Surgery* 69:554–56
71. Ribot, S., Ruthfeld, D., Frankel, H. J. 1972. Infectious endocarditis in maintenance hemodialysis patients: a report of four episodes among three patients. *Am. J. Med. Sci.* 264:183–88
72. Lillihei, L. C., Robb, J. R. R., Visscher, M. B. 1950. The occurrence of endocarditis with valvular deformities in dogs with arteriovenous fistulae. *Ann. Surg.* 132:577–90
73. Lillihei, L. C., Shaffer, J. M., Spink, W. W., Robb, J. R. R., Wargo, J. D., Visscher, M. B. 1951. Role of cardiovascular stress in the pathogenesis of endocarditis and glomerulonephritis. *Arch. Surg.* 63:421–34
74. Gowdy, R. A., Martin, S. P., Kerby, G. P., Sealy, W. C. 1952. Pathogenesis of valvular endocarditis. *Arch. Surg.* 65:271–75
75. Kaye, D. 1972. Bacterial endocarditis in the presence of arteriovenous fistulae. *Am. J. Med. Sci.* 264:189–90
76. Sande, M. A., Levison, M. E., Lukas, D. S., Kaye, D. 1969. Bacteremia associated with cardiac catheterization. *N. Engl. J. Med.* 281:1104–6
77. Patton, R. D., Kenamore, B., Stein, E. 1969. Antibiotic prophylaxis for temporary pacemakers. *N. Engl. J. Med.* 281:1106–8
78. Corman, L. C., Levison, M. E. 1975. Sustained bacteremia and transvenous cardiac pacemakers. *J. Am. Med. Assoc.* 233:264–66
79. Stamm, W. E., Colella, J. J., Anderson, R. L., Dixon, R. E. 1975. Indwelling arterial catheters as a source of nosocomial bacteremia: an outbreak caused by Flavobacterium species. *N. Engl. J. Med.* 292:1099–1102
80. Transducer-associated bacteremia. 1975. *Morb. Mortal. Wkly. Rep.* 24:295
81. Yeh, T. J., Anabtawi, I. N., Cornett, V. E., White, A., Stern, W. H., Ellison, R. G. 1967. Bacterial endocarditis following open-heart surgery. *Ann. Thorac. Surg.* 3:29–36
82. Killen, D. A., Collins, H. A., Koenig, G. A., Goodman, J. S. 1970. Prosthetic cardiac valves and bacterial endocarditis. *Ann. Thorac. Surg.* 9:238–47
83. Blakemore, W. S., McGarrity, G. S., Thurer, R. J,, Wallace, H. W., MacVaugh, H. III, Coriell, L. L. 1971. Infection by airborne bacteria with cardiopulmonary bypass. *Surgery* 70:830–38
84. Ankeney, J. L., Parker, R. F. 1969. Staphylococcal endocarditis following open-heart surgery related to positive intraoperative blood cultures. In *Prosthetic Heart Valves,* ed. L. A. Brewer III, pp. 719–30. Springfield, Ill: Thomas
85. Johnson, W. D., Cobbs, G. C., Arditi, L. I., Kaye, D. 1968. Diphtheroid endocarditis after insertion of a prosthetic heart valve. *J. Am. Med. Assoc.* 203:117–19
86. McKittrick, L. S., Wheelock, F. C. 1959. The routine use of antibiotics in elective abdominal surgery. *Surg. Gynecol. Obstet.* 99:376–77
87. Kaye, D. 1976. *Infective Endocarditis.* Baltimore: Univ. Park Press. 272 pp.
88. Slaughter, L., Morris, J. E., Starr, A. 1973. Prosthetic valve endocarditis: a 12 year review. *Circulation* 47:1319–26
89. Dismukes, W. E., Karchmer, A. W., Buckley, M. J., Austen, W. G., Swartz, M. N. 1973. Prosthetic valve endocarditis: analysis of 38 cases. *Circulation* 48:365–77
90. Wilson, W. R., Jaumin, P. M., Danielson, G. E., Giuliani, E. R., Washington, J. A. II, Geraci, J. E. 1975. Prosthetic valve endocarditis. *Ann. Intern. Med.* 82:751–56
91. Geraci, J. E., Dale, A. J. D., McGoon, D. C. 1963. Bacterial endocarditis and endarteritis following cardiac operations. *Wis. Med. J.* 62:302–15
92. Nelson, R. M., Jenson, L. B., Peterson, C. A., Sanders, B. C. 1965. Effective use of prophylactic antibiotics in open-heart surgery. *Arch. Surg.* 90:731–36
93. Amoury, R. A., Bowman, F. O. Jr., Malm, J. R. 1966. Endocarditis associated with intracardiac prostheses: diagnosis, management and prophylaxis. *J. Thorac. Cardiovasc. Surg.* 51:36–47
94. Stein, P. D., Harken, D. W., Dexter, L. 1966. The nature and prevention of prosthetic valve endocarditis. *Am. Heart J.* 71:393–407
95. Sande, M. A., Johnson, W. D., Hook, E. W., Kaye, D. 1972. Sustained bacteremia in patients with prosthetic cardiac valves. *N. Engl. J. Med.* 286: 1067–70
96. Fekety, F. R., Cluff, L. E., Sabiston, D. E., Seidl, L. G., Smith, J. W., Thoburn, R. 1969. A study of antibiotic prophylaxis in cardiac surgery. *J. Thorac. Cardiovasc. Surg.* 57:757–63
97. Goodman, J. S., Schaffner, W., Collins, H. A., Battersby, E. J., Koenig, M. G.

1968. Infection after cardiovascular surgery: clinical study including examination of antimicrobial prophylaxis. *N. Engl. J. Med.* 278:117–23

98. Conte, J. E., Cohen, S. N., Roe, B. B., Elashoff, R. M. 1972. Antibiotic prophylaxis and cardiac surgery: a prospective double-blind comparison of single-dose versus multiple-dose regimens. *Ann. Intern. Med.* 76:943–49
99. Weinstein, L. 1972. Infected prosthetic valves: a diagnostic and therapeutic dilemma. *N. Engl. J. Med.* 286:1108–9
100. Weinstein, L., Goldfield, M., Chang, T. 1954. Infections occurring during chemotherapy. *N. Engl. J. Med.* 251: 247–55
101. Seelig, M. S., Speth, C. P., Kozinn, P. J., Taschdjian, C. L., Foni, E. F., Goldberg, P. 1974. Patterns of *Candida* endocarditis following cardiac surgery: importance of early diagnosis and therapy. *Prog. Cardiovasc. Dis.* 17: 125–60
102. Carrothers, M. M., Kanokvechayant, R. 1973. *Pseudomonas aeruginosa* endocarditis: report of a case, with review of the literature. *Am. J. Med.* 55:811–18
103. Tuazon, C. U., Hill, R., Sheagren, J. N. 1974. Microbiologic study of street heroin and injection paraphernalia. *J. Infect. Dis.* 129:327–29
104. Tuazon, C. U., Sheagren, J. N. 1975. Staphylococcal endocarditis in parenteral drug abusers: source of the organism. *Ann. Intern. Med.* 82:788–90
105. Maki, D. G., Goldmann, D. A., Rhame, F. S. 1973. Infection control in intravenous therapy. *Ann. Intern. Med.* 79:867–87
106. Goldmann, D. A., Maki, D. G., Rhame, F. S., Kaiser, A. B., Tenney, J. H., Bennett, J. V. 1973. Guidelines for infection control in intravenous therapy. *Ann. Intern. Med.* 79:848–50
107. Buchholz, D. H., Young, U. M., Friedman, N. R., Reilly, J. A., Mardiney, M. R. Jr. 1971. Bacterial proliferation in platelet products stored at room temperature. *N. Engl. J. Med.* 285:429–33
108. Hockett, R. N., Beers, R. L., Loesche, W. L. 1972. Low-level bacteremia in humans. *Abstr. Ann. Meet. Am. Soc. Microbiol.* M6, p. 81
109. Reith, A. F., Squier, T. I. 1932. Blood cultures in apparently healthy persons. *J. Infect. Dis.* 51:336–43
110. Elliott, S. D. 1939. Bacteremia and oral sepsis. *Proc. R. Soc. Med.* 32:747–54
111. Elliott, S. D. 1939. Bacteremia following tonsillectomy. *Lancet* 2:589–92
112. Austrian, R., Gold, G. 1964. Pneumococcal bacteremia with especial reference to bacteremic pneumococcal pneumonia. *Ann. Intern. Med.* 60:759–76
113. Cherubin, L. E., Neu, H. C. 1971. Infective endocarditis at the Presbyterian Hospital in New York City from 1938 to 1967. *Am. J. Med.* 51:83–96
114. Shinebourne, E. A., Cripps, C. M., Hayward, G. W., Shooter, R. A. 1969. Bacterial endocarditis, 1956–1965: analysis of clinical features and treatment in relation to prognosis and mortality. *Br. Heart J.* 31:536–42
115. Gelfman, R., Levine, S. A. 1942. The incidence of acute and subacute bacterial endocarditis in congenital heart disease. *Am. J. Med. Sci.* 204:324–33
116. Uwaydah, M. M., Weinberg, A. N. 1965. Bacterial endocarditis—a changing pattern. *N. Engl. J. Med.* 273: 1231–35
117. Roberts, W. C., Perloff, J. K., Costantino, T. 1971. Severe valvular aortic stenosis in patients over 65 years of age. *Am. J. Cardiol.* 27:497–506
118. Pomerance, A. 1972. Pathogenesis of aortic stenosis and its relation to age. *Br. Heart J.* 34:569–74
119. Burnside, J. W., DeSanctis, R. W. 1972. Bacterial endocarditis on calcification of the mitral anulus fibrosus. *Ann. Intern. Med.* 76:615–18
120. Joffe, S., Feil, H. 1955. Subacute bacterial endocarditis arising in mural thrombi following a myocardial infarction: a case report. *Circulation* 12: 242–26
121. Bernstein, W. H., Robinson, M. J., Baer, B., Cohen, L. S., Samet, P. 1969. Postmyocardial infarction septal perforation complicated by bacterial endocarditis. *Am. J. Cardiol.* 24:432–35
122. Vecht, R. J., Oakley, C. M. 1968. Infective endocarditis in three patients with hypertrophic obstructive cardiomyopathy. *Br. Med. J.* 2:455–59
123. Cardelia, J. U., Befeler, B., Hildner, F. J., Samet, P. 1971. Hypertrophic subaortic stenosis complicated by aortic insufficiency and subacute bacterial endocarditis. *Am. Heart J.* 81:543–47
124. Jeresaty, R. M. 1973. Mitral valve prolapse-click syndrome. *Prog. Cardiovasc. Dis.* 15:623–51
125. Soman, V. R., Breton, G., Hershkowitz, M., Mark, H. 1974. Bacterial endocarditis of mitral valve in Marfan's syndrome. *Br. Heart J.* 36:1247–50
126. Morgan, W. L., Bland, E. F. 1959. Bac-

terial endocarditis in the antibiotic era. *Circulation* 19:753–65

127. Pankey, G. A. 1961. Subacute bacterial endocarditis at the University of Minnesota Hospital, 1939 through 1959. *Ann. Intern. Med.* 55:550–61
128. Garrison, P. K., Freedman, L. R. 1970. Experimental endocarditis I: staphylococcal endocarditis resulting from placement of a polyethylene catheter in the right side of the heart. *Yale J. Biol. Med.* 42:394
129. Durack, D. T., Beeson, P. B., Petersdorf, R. G. 1973. Experimental bacterial endocarditis III: production and progress of the disease in rabbits. *Br. J. Exp. Pathol.* 54:142–51
130. Valone, J. A., Freedman, L. 1976. Studies in experimental endocarditis. *J. Infect. Dis.* In press
131. Rogers, D. E. 1958. Staphylococcal infections. In *Disease-a-Month,* pp. 22–24. Chicago: Yearbook
132. Durack, D. T., Petersdorf, R. G. 1973. Chemotherapy of experimental streptococcal endocarditis I: comparison of commonly recommended prophylactic regimens. *J. Clin. Invest.* 52:592–98
133. Pelletier, L. L., Durack, D. T., Petersdorf, R. G. 1975. Chemotherapy of experimental streptococcal endocarditis IV: further observations on prophylaxis. *J. Clin. Invest.* 56:319–30
134. Bender, I. B., Selzer, S., Yermish, M. 1960. The incidence of bacteremia in endodontic manipulation. *Oral Surg.* 13:353–60

Ann. Rev. Med. 1977. 28:393–409
Copyright © 1977 by Annual Reviews Inc. All rights reserved

THE ROLE OF ETHANOL IN CARDIAC DISEASE[1]

❖7251

Timothy J. Regan, M.D., Philip O. Ettinger, M.D., Bunyad Haider, M.D., S. Sultan Ahmed, M.D., Henry A. Oldewurtel, and Michael M. Lyons, M.D.
Department of Medicine, College of Medicine and Dentistry of New Jersey,
New Jersey Medical School, Newark, New Jersey 07103

INTRODUCTION

Acceptance of the concept of alcoholic cardiomyopathy has varied over the years. In Europe during the last century, there were repeated reports of diffuse disease of heart muscle in chronic alcoholic individuals, who apparently had low-output heart failure (1). Periods of economic distress with attendant nutritional deprivation have resulted in a different clinical presentation, the high-output state of beriberi due to thiamine deficiency. Some concluded that this was the only cardiovascular effect of chronic alcoholism. However, in recent decades when the industrialized countries have had a generally sufficient foodstuff supply, there have been numerous reports of low-output heart failure in alcoholics, without evidence of nutritional deficiency in the majority (2–4).

A recent nutritional survey of alcoholics who were generally considered not to be indigent revealed a relatively low incidence of nutritional deficiency, although 36% of their calories were derived from alcohol (5). Only 15% of these patients had thiamine deficiency as determined by diet history and excretion of vitamin metabolites. Riboflavin, niacin, and protein deficiency occurred in less than 10% of this group, and low levels of serum albumin were observed in an even smaller segment of the group. Whether the myocardial response to chronic alcoholism differs in the minority of patients with nutritional deficiency, when compared to the better-nourished alcoholic, remains to be evaluated. It should be noted that blood vitamin levels assayed in hospitalized patients frequently reveal no clear distinction between alcoholics with and without evidence of liver disease (6). Blood vitamin concentrations may also be low as a nonspecific effect of chronic illness in the absence of alcoholism.

---

[1] This investigation was supported in part by Research Grant No. AA00242 from the National Institute on Alcohol Abuse and Alcoholism and Postgraduate Training Grant No. HL 05510 from the National Heart and Lung Institute.

In considering the influence of ethyl alcohol on the heart, it is pertinent to reflect on the prevalence of alcoholism in the general population. In view of the difficulty of obtaining a reliable history in many subjects with this disease, we can assume that the reported incidence of alcoholism is likely to be underestimated. In a recent survey, Cahalan and his associates found that two thirds of the population used alcohol, and were categorized in four groups: infrequent, light, moderate, and heavy users (7). Heavy users represented about 12% of the adult population. An important element in considering prevalence is the variation between different regions of the country and between hospitals of a given region.

RESPONSE TO ACUTE ETHANOL USE

A number of investigations have indicated that ethyl alcohol may simultaneously affect the function of many organs. Thus, in contrast to the popular view that ethanol has beneficial effects on the heart, doses that are mildly intoxicating have been shown to adversely affect left-ventricular function in some circumstances. This effect is in part dependent on prior experience with alcohol; that is, larger doses are required to demonstrate impaired pumping action of the heart in the chronic alcoholic subject who has no clinical evidence of heart disease (8). However 6 oz. of scotch, fed to normal individuals over a two-hour period, has been found to diminish the force of heart-muscle contraction at a mildly intoxicating blood level of 75 mg/100 ml. This effect progresses as blood levels rise. It is usually dissipated within a few hours after the drinking is interrupted (9). Supporting the view that ethanol is acutely depressant to myocardium is the fact that ventricular dysfunction is rapidly reversed after 15–30 min of hemodialysis (10).

In contrast to these findings in alcoholics without clinical evidence of heart disease, the patient who has already had at least one episode of heart failure may exhibit a greater sensitivity to the 6-oz dose of scotch (11) with substantial elevation of filling pressure. The response in cardiac subjects who are not alcoholics may be qualitatively similar (12). An important variation of the acute response to ethanol is the combination with other pharmacologic agents. Accidental or suicidal deaths have resulted from the combination with barbiturates. Potentiation of the myocardial depressant effect has been observed experimentally with such a combination (13), which is concurrent with respiratory depression.

The sympathetic nervous system presumably has an important role in modifying the cardiac response. Thus, blockade of the system has been shown to produce greater depression of left-ventricular function during ethanol administration (14). Intensification of the cardiovascular response to ethanol in animals treated with disulfiram, an inhibitor of beta hydroxylation with resultant lower levels of endogenous norepinephrine (15), may well be on a similar basis (16). It is noteworthy that the adrenal gland does not appear to respond with greater secretion of adrenaline or noradrenaline, at least in response to moderately intoxicating blood levels (17), so that cardiac catecholamine stores would appear to be the major factor modifying the response to acute ethanol intake.

In considering the direct influence of ethanol on cardiac cells when administered acutely, changes of salts in the muscle are of major importance. The action in reducing force of contraction may well entail altered calcium movement in the cell (18). It is known that potassium and phosphate transiently leak out of the muscle cells after a 12-oz dose (8), an effect that is not attributable to coronary blood flow reduction. The loss of cations may be related to an inhibitory effect of ethanol on active transport of potassium and sodium across cell membranes, which has been suggested as a basic mechanism of action of ethanol upon most cells (19).

Another metabolic change is represented by alteration of lipid transport in the myocardium: the large dose of ethanol reduced the uptake of free fatty acid by the left ventricle, while the increased triglyceride uptake resulted in accumulation of lipid in myocardium (20). This response may contribute to pathologic changes observed in human myocardium at postmortem examination; substantial increases in lipids, presumably triglycerides, have been observed in the alcoholic heart (21).

It is noteworthy that noncardiac striated muscle, including uterine (22) and skeletal muscle (23), are also depressed acutely by ethanol, although in the diaphragm this follows a period of enhanced contractility. With chronic use, the asymptomatic alcoholic may exhibit modest reduction of lactic dehydrogenase, as well as a small reduction in fast-twitch glycolytic fibers and volume of mitochondria (24). The conditions required for acute myopathy and rhabdomyolysis are not well defined, but malnutrition and hypophosphatemia are frequently associated (25).

CHRONIC ETHANOL USE IN ANIMAL MODELS

While histologic abnormalities have been observed in 90% of unselected alcoholics in one postmortem series (26), uncertainty as to the quantity of ethanol intake, the nutritional status of the patient, including electrolyte deficits and the possibility of heart disease from other causes, have obscured the relationship of excessive ethanol use and cardiomyopathy. Previous chronic experiments with ethanol have yielded conflicting data in terms of the production of a functional deficit (27, 28), but the negative study used smaller quantities and was relatively short-term.

To eliminate some of the above variables, a group of young adult male dogs was maintained in a relatively normal nutritional state while receiving up to 36% of calories as ethanol, approximating the quantity reported in a population of human alcoholics (5). After an average of 18 months they were found to have maintained weight, hematocrit, serum proteins, vitamins, and electrolytes to a similar extent as the normal controls. They were anesthetized to permit determination of left-ventricular performance, conduction, morphology, and metabolism using methods previously described (29).

Cardiac Function

Mean heart rate and aortic pressure were similar in both groups. However, in the chronic-ethanol animals there was a significantly higher end-diastolic pressure despite a lower level of end-diastolic volume. Assessment of left ventricular function was undertaken during afterload increments with angiotensin of similar extent in both groups. Stroke output increased moderately but significantly in the controls, while no such increments occurred in the alcoholic animals. Ejection fraction was similar in both groups before and after angiotensin. A rise of stroke work in the normals from 1.39 ± 0.23 to 1.81 ± 0.21 g m/kg was significantly less in the experimental group ($P < 0.01$). Normal controls exhibited a moderate rise of end-diastolic pressure and end-diastolic volume during angiotensin infusion. However, a significantly larger rise of end-diastolic pressure occurred in the ethanol group, while the end-diastolic volume response was significantly less than in controls. This is presumably the basis for the reduced stroke volume.

To further analyze the apparent altered stiffness of the ventricle, six of the normal animals and six of the ethanol group were infused with normal saline via a catheter in the left ventricle. At an infusion rate of 50 ml/min for 3–4 min, the normals generally exhibited a proportionate rise of left-ventricular end-diastolic pressure and volume. In the ethanol group, there was a significantly higher rise of end-diastolic pressure than in the normal, despite an elevation of end-diastolic volume proportionate to that of controls. Heart rate declined and stroke volume rose similarly in both groups without a change in afterload, so that the significantly higher end-diastolic pressure in the experimental group does not appear to be related to these variables. A potential morphologic basis for the altered diastolic pressure-volume relations was found in sections of the left ventricle. Staining with Alcian blue showed distinct accumulation of a glycoprotein-like material in the interstitium of the left ventricular wall in the ethanol group. The degree of staining ranged from 2 to 3+ on a scale of 0 to 4. There was virtually no staining in the normal controls.

As a further assessment of the myocardium in alcoholic animals, high-speed, high-frequency electrocardiograms were obtained. QRS time was distinctly abnormal in the longer-term animals (Figure 1) (30). Similarly, the conduction time from the His bundle to the onset of ventricular depolarization was significantly prolonged. Dilatation of the intercalated disc by electron microscopy appeared to be the morphologic correlate (Figure 2).

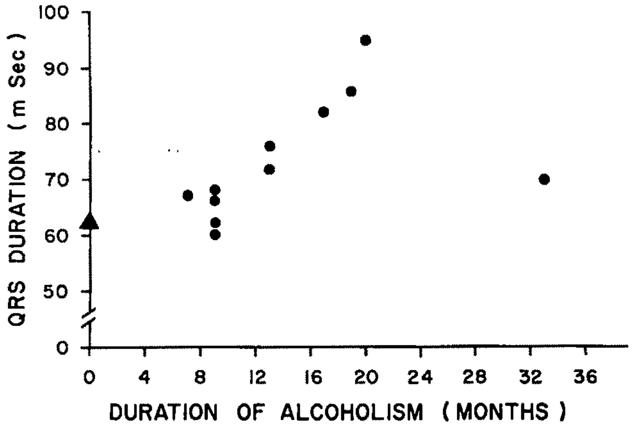


Figure 1 The conduction time in the ventricular wall is delayed to a greater extent in the longer-term alcoholic animals. (Reprinted by permission from *Am. Heart J.*)

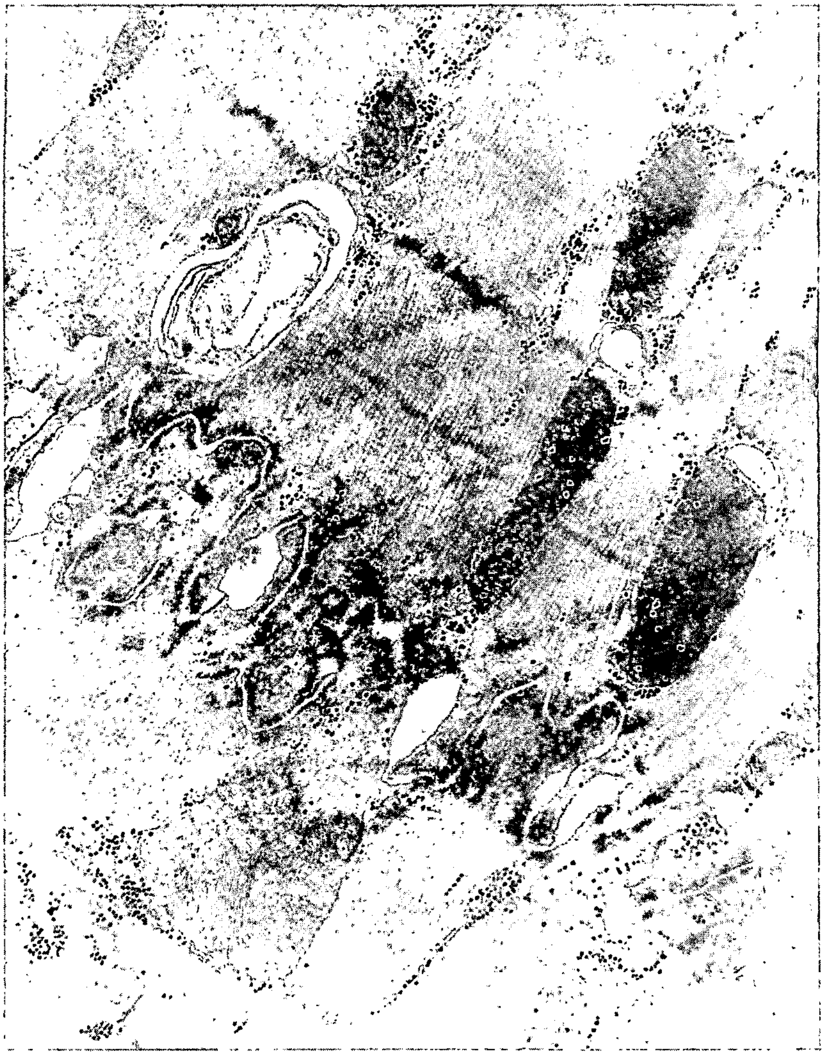


Figure 2 Intercalated disc of alcoholic dog. Cystic areas of dilatation of undifferentiated regions are apparent. The nexus region appears normal (Magnification 9500 X, 47,500 X final). (Reprinted by permission from *Am. Heart J.*)

Myocardial Metabolism

Analyses of left-ventricular lipids revealed an increment of triglyceride in the ethanol group (29), which appeared as fine cytoplasmic droplets on Oil Red O stain. Cholesterol, phospholipid, and free fatty-acid levels were not significantly altered in the three layers of myocardium compared to those of the control group, while plasma lipid concentrations were normal.

To explore the metabolic basis for these lipid alterations, ^{14}C–1-oleic acid was administered by continuous infusion into the inferior vena cava. In the chronic-alcohol state no significant uptake of triglyceride was observed and free fatty-acid uptake did not differ significantly from controls, but $^{14}CO_2$ production was decreased. Reduced oxidation may be secondary to diminished activity of rate-limiting enzymes in the citric-acid cycle, demonstrated after chronic ethanol use (31). A resultant enhanced availability of acyl-CoA, perhaps associated with enhanced glycerol acyltransferase activity (32), may account for triglyceride accumulation.

Incorporation of labeled fatty acid into phospholipid was sufficiently reduced in subendocardium to diminish specific activity, despite the observation that the myocardial (^{14}C) free fatty-acid fraction was not reduced. This may represent interference with transacylation in the phospholipid molecule. A recent report of diminished arachidonate in cardiac phospholipids of the ethanol-treated rat supports the view that substantial fatty-acid compositional changes occur in this lipid class after chronic alcohol use (33).

A previous hemodynamic study of chronic ethanol administration in the rat has shown a significant decrease in "potential ventricular force" (27). In contrast, a definite abnormality of ventricular function was not found in the same species fed less ethanol, 15% by volume (28), compared to 25% in the previous series. An important methodologic difference existed, however, in that ventricular function was not tested at the uppermost segment of the ascending portion of the function curve. Since caloric equivalents were not provided, it is difficult to compare these data with the present study; moreover, species differences may exist. In a recent study of dogs fed ethanol in a daily quantity similar to that of the present study, but for a 14-week period, no hemodynamic abnormality was observed (31). This is consistent with the view that duration of ingestion is an important determinant.

The importance of chronicity is also evident in the reported failure to find morphologic abnormalities of cardiac cell organelles in animals ingesting ethanol for 1–3 months (34, 35). Dilatation of the sarcoplasmic reticulum and intercalated disk were the only unequivocal structural abnormalities in the relatively well-nourished animals in this current report, findings also observed in two studies of patients with alcoholic cardiomyopathy (36, 37). It is problematic whether the extensive abnormalities of myofibrils and mitochondria reported in other human studies are secondary to multiple episodes of heart failure or are due solely to toxicity, coexistent nutritional deficiency, cobalt excess, or other complications (38–40).

Although with time most species appear to develop myocardial abnormalities, the duration and quantity required to develop functional or morphologic alterations appears to vary. Thus the rhesus monkey, surprisingly, has shown focal myocytolysis and early fibrosis after three months of ingesting ethanol as 40% of total calories

(41). It is possible that the feeding modality, relatively rapid by stomach tube, or the synthetic liquid diet contributed to this response. In the rat, ethanol as 10% of total calories has apparently produced ultrastructural abnormalities as early as seven weeks (42). Prior experience with rodents and their proclivity for viral infections (43) raises the question of possible conditioning of the myocardium, with sensitization to the early development of lesions at relatively low doses.

Pathogenesis of Malfunction

At this stage of alcoholism in the dog, there was no hypertrophy, the coronary arteries were normal, and morphologic studies revealed no inflammatory response in heart muscle to complicate the assessment of chemical composition. The changes in lipid metabolism were not accompanied by unequivocal abnormalities of mitochondrial structure on electron micrographs, but accumulation of glycogen-type particles was evident. Lysosomal structures were relatively infrequent in both groups. Dilatation of the sarcoplasmic reticulum was a prominent feature in all animals receiving ethanol. Of particular interest in view of the observed prolonged conduction times was the finding that the undifferentiated portion of the intercalated disc was similarly affected.

Conduction delay might also be effected by altered cation distribution. The ethanol group exhibited a significant transmural gradient of potassium ion, with reduction of approximately 10% in the endocardium compared to controls. The basis for the altered cation content is not known, but dysfunction of cell membrane on the basis of the change in phospholipid composition may be present.

On the basis of the long-term observations in the canine model (29), an altered diastolic pressure-volume relationship appears to be based on accumulation of Alcian-positive glycoprotein in the interstitium. The factors responsible for progression to interstitial collagen accumulation as seen in man are not known. This may limit the rapid accessibility of calcium to intracellular sites and ultimately contribute to the development of heart failure. Whether independent abnormalities of intracellular calcium transport or myosin ATPase are present in long-term experiments has yet to be demonstrated. A difficulty in assessing the importance of these factors is the fact that no animal model has yet been of sufficient duration to produce frank heart failure.

In considering the agent of injury in addition to ethanol, its first metabolite acetaldehyde can significantly depress left-ventricular function (44) when infused at rates that simulate blood levels associated with moderate ethanol ingestion. This suggests that acetaldehyde may have a role in the pathogenesis of the cardiomyopathy. Although there is no evidence that alcohol dehydrogenase is induced in the myocardium during chronic alcoholism, such an adaptation would presumably increase the concentration of acetaldehyde in the cardiac cells and enhance the potential for cell injury. An additional consideration involves the role of magnesium ion (45). Primary magnesium deficiency does produce cardiac lesions in immature experimental animals. Unequivocal disease in adults is difficult to demonstrate, however, although secondary potassium loss could have important consequences. As a general observation, functional and clinical abnormalities of myocardium frequently occur with normal serum levels of this cation.

PRECLINICAL MALFUNCTION IN MAN

To determine whether there is an equivalent functional abnormality in man, studies were undertaken of alcoholic subjects with no symptoms or clinical evidence of heart disease or malnutrition (8). Documentation of 10–15 years of alcoholism and the type of ethanol used was obtained from the patients' histories or from close relatives. Whiskey was the predominant alcoholic beverage consumed several times a week in amounts ranging from 0.5–2 pints/day. Liver biopsy revealed fatty liver without fibrosis. Left-ventricular performance was studied by increasing afterload with angiotensin. The noncardiac alcoholic exhibited a significantly greater rise of ventricular end-diastolic pressure with a minimal increment of stroke volume compared to nonalcoholic controls, confirmed in a subsequent study (46). Diminished ventricular performance has also been observed during exercise in cirrhotic patients without clear evidence of cardiac disease (47). Noninvasive studies that measure systolic time intervals have also confirmed that many asymptomatic subjects have modest depression of left-ventricular function (48, 49). It is noteworthy that noncardiac alcoholics have been shown to accumulate Alcian-positive material in the myocardial interstitium (Figure 3) (50), which may be a basis for the functional abnormality, as postulated in the canine studies (29). Some degree of interstitial fibrosis may be present at this stage, as suggested by postmortem studies of accident victims who were alcoholic (51).

CARDIAC FAILURE

To test the thesis that cumulative effects of ethanol over a period of time may result in cardiac abnormality despite adequate nutrition, a well-compensated patient was fed Scotch whiskey for a period of 5½ months at a daily dose of 12–16 oz (8). After an interval of 6 weeks (Figure 4), resting heart rate began to rise and there was prolongation of circulation time and elevation of venous pressure without evidence of malnutrition. After 4 months, a ventricular gallop rhythm appeared, which persisted until ethanol was interrupted. Subsequently, without specific cardiac therapy, there was spontaneous restoration to normal. A major role of ethanol in the production of left heart failure in this subject was substantiated by the gradual reversion of the cardiocirculatory abnormality after alcohol ingestion was interrupted. This observation supports the thesis that the myocardial disease is reversible at certain stages if intake of ethanol is discontinued.

In circumstances marked by progression of cardiac dysfunction, pulmonary congestion may lead to exertional or nocturnal dyspnea. If long sustained, or after repeated episodes, pulmonary hypertension and right heart failure may become evident. Unless there is complicating papillary muscle insufficiency giving rise to mitral regurgitation, cardiomegaly may be moderate in extent and heart size may revert to near normal after central congestion is corrected. In some patients, systemic or pulmonary emboli from mural thrombi may adversely effect the patient's course. These clinical episodes frequently seem to be precipitated by intensified drinking episodes.

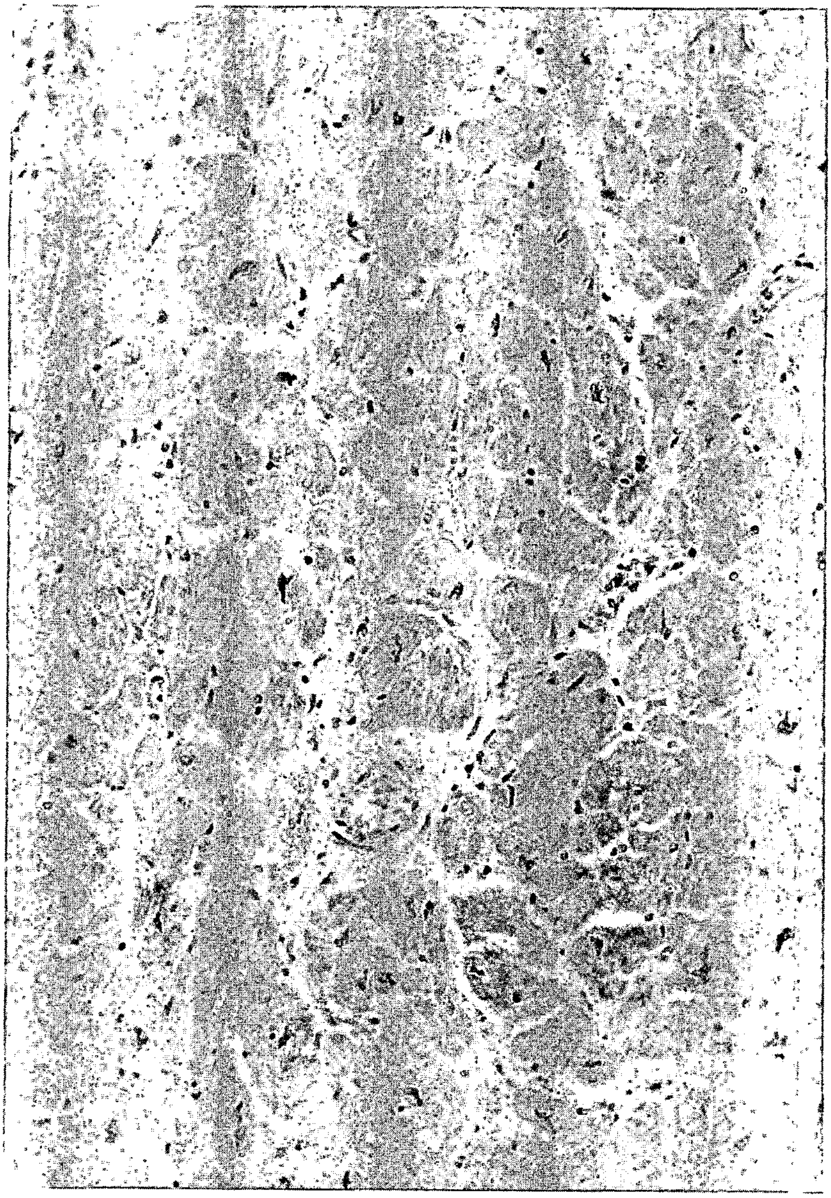


Figure 3 Alcian blue–positive glycoprotein accumulated in left-ventricular interstitium between muscle fibers. None was present in control. (Reprinted by permission from *Circulation.*)

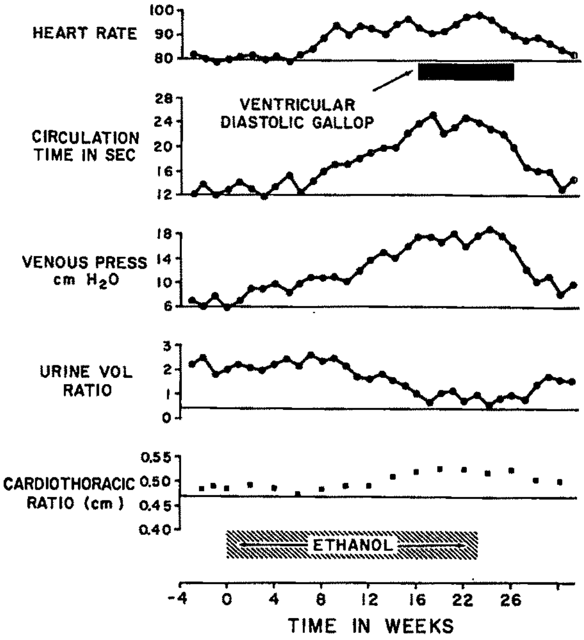


Figure 4 These are observations in a single, well-nourished patient receiving daily Scotch, which resulted in evidence of heart failure. The failure regressed without medical treatment after interrupting alcohol intake. (Reprinted by permission from *J. Clin. Invest.*)

Several reports of cardiomyopathy have emphasized the difficulty in obtaining a history of alcoholism (2–4). There is a male predominance, and suggestive diagnostic aspects include social disruption, accident-proneness, and a familial history. The major positive diagnostic feature is the history of ingesting ethanol in intoxicating amounts for many years, frequently with spree episodes. This information may often be obtained only through persistent questioning of the patient or from relatives. Negative aspects include exclusion of other causes of heart disease, namely hypertension, coronary-artery disease, cor pulmonale, and congenital or valvular disease. Other forms of cardiomyopathy (viral, infiltrative, metabolic, etc) must also be considered. In addition to the usual lack of specificity of the clinical signs and symptoms, high diastolic arterial blood pressure is not infrequently found during periods of severe congestive heart failure, which may lead to the false diagnosis of hypertensive heart disease. The blood pressure returns to normal following response to therapy. Hepatic cirrhosis and clear evidence of peripheral neuritis are not frequently present in patients with alcoholic cardiomyopathy.

Not only are quantity, frequency, and duration of ethanol intake difficult to assess, but their relationship to organ pathology is not straightforward. It is probable that these factors have a similar importance for the varied chronic disease secondary to alcohol abuse, so that information in relation to liver disease for which there is more complete data may in a general way be applicable to the heart. Alcoholic hepatitis usually develops after years of excessive drinking, although in a few patients it appears within a few years, but it may not develop in many even after several decades of drinking (52). Over 80% were drinking five years or longer before developing symptoms. The probability of developing alcoholic hepatitis is small in those who drink less than 80 g/day of ethanol (approximately 8 oz of 86-proof whiskey or 1 liter of wine) (53–55), or if ethanol provides less than 20% of total calories (56). As the daily ethanol consumption increases to 80–160 g/day and the duration of drinking becomes longer, the risk of developing alcoholic liver disease increases (57, 58). The probability of alcoholic hepatitis or cirrhosis is great if the daily ethanol consumption exceeds 160 g (for example, two thirds of a bottle of whiskey, the approximate daily metabolic capacity of an average nondrinking person), and if the drinking—either steady or in sprees—persists for 10–15 years or longer (56–59). It is noteworthy that the rate of ethanol metabolism can be high enough to permit the consumption of large amounts without raising the blood ethanol level over 80 mg/100 ml, the concentration at which the conventional breath analyzer begins to detect ethanol (60).

Atypical Myocardial Infarction

Precordial pain is not infrequently observed in myopathies of varied etiology; however, the association of alcoholism with the clinical syndrome of acute myocardial infarction has not previously been observed. Transmural myocardial scar has been found at postmortem examination in subjects with alcoholism and without significant coronary atherosclerosis. To elucidate this realtionship, alcoholic subjects admitted to a coronary-care unit (CCU) with unequivocal evidence of acute transmural myocardial infarction were investigated (50). Ten of twelve were without traditional coronary risk factors. Those with prior ECG's frequently exhibited absent Q waves in Leads I, V_5, and V_6, which became manifest as the infarction evolved. Examination of the coronary arteries at postmortem or by angiography revealed no significant occlusive lesions.

Morphologic examination of the myocardium revealed concentric periarterial fibrosis, which was postulated to restrict coronary-flow increments during periods of high–blood flow requirements. This phenomenon is considered analogous to cardiac-muscle necrosis associated with the periarterial lesions of constrictive pericarditis, perhaps conditioned by the abnormal metabolism of cardiac cells in chronic alcoholism.

A thromboembolic process was deemed unlikely in this study on the basis of several observations. The patients without previous infarction did not have prior heart failure, cardiomegaly, evidence of arrhythmias, or valvular disease, which appear to be prerequisites for embolization from the left heart chambers (61). Thrombocytosis and drugs affecting platelets were also not present. These patients were considered to have a preclinical form of toxic heart-muscle disease prior to infarction and mural thrombi have not been described at this stage.

Crucial to the thromboembolic postulate is the demonstration in such patients that emboli emanating from the left heart are distributed to the systemic circulation in an incidence approximating that of proven mural thrombi. As a clinical model for the distribution of emboli in the arterial circulation, patients with prosthetic valves and a relatively high incidence of embolization show a striking noncoronary preponderance (62). None of the total of 12 patients in this series, representing 20% of CCU admissions for the given period, had evidence of embolization to any organ. Since angiography was not feasible during acute infarction, the presence of thromboembolism was assessed indirectly by serial determinations of circulating platelet levels, which are known to be altered during systemic thromboembolic phenomena (63, 64). The absence of a change in platelet count during the first day suggested that accelerated platelet turnover as a reflection of arterial thromboembolism was unlikely in these patients.

ECG, CONDUCTION, AND ARRHYTHMIAS

Occasionally, palpitation is a major complaint in alcoholic cardiomyopathy and is presumably due to cardiac arrhythmias. A recent report indicates that atrial arrhythmias may occur in alcoholics who have not experienced heart failure, and present with palpitation or syncope (65), without recurrence unless heavy ethanol ingestion is resumed. The arrhythmia often regresses spontaneously. Sinus tachycardia associated with ventricular premature beats has been suggested as an important clue to the diagnosis of alcoholic cardiomyopathy, but this combination is common in other forms of heart disease. On the ECG, abnormalities of repolarization are the rule, but the spinous or cloven-shaped T waves originally described by Evans (66) are found infrequently. Furthermore, these changes appear to be related to cardiac-muscle damage rather than to a specific cause. Whether or not there are early electrocardiographic abnormalities that antedate clinical disease is obviously of interest. Levine and his co-workers concluded that the sharpness of T waves was the only possible electrocardiographic clue to alcoholism (67). The absence of septal Q waves in the alcoholic has been suggested as grounds for considering the diagnosis of cardiomyopathy (68). Gross conduction changes in the form of AV block and left- or right-bundle branch block occur in a minority of patients.

It is important to be aware that sudden unexpected death, thought to be due to ventricular fibrillation, has been observed in a relatively high incidence in a group of young adult alcoholics (69). These subjects had fatty livers as the principal pathologic finding. Additional information in this regard is derived from a study of outpatient alcoholics in Oslo, in whom an increased incidence of sudden cardiac death was observed (70). Experimental evidence of a reduced fibrillation threshold in a chronic-ethanol animal model (71) provides further support for this phenomenon.

TREATMENT AND REVERSIBILITY

Management of alcoholic cardiomyopathy depends on the state of the disease when the addicted individual is first seen. Even in the absence of cardiac failure or enlargement, patients with unexplained arrhythmias or absent septal Q waves should raise a high index of suspicion requiring verification of a negative history for alcoholism. The key to treatment at all stages involves complete abstinence. The alcoholic with cardiac abnormality may be approached effectively on an individual physician-patient relation as well as in group therapy.

In the only available study of its kind (72), almost one third of patients followed on an outpatient basis were found to have apparently maintained abstinence over a three-year period (Table 1). The majority of these had an improved or unchanged cardiac status. Deaths occurred preponderantly in those who did not abstain. Some uncertainties exist, since the relative cardiac status on admission to the study was not described. It is noteworthy that over 20% of those who were allegedly abstinent deteriorated in cardiac status. Presumably, at certain stages of the disease the pathogenetic mechanisms may continue unabated.

After the onset of clinical manifestations, traditional antiarrhythmic agents, dc countershock, digitalis, and diuretics may be used as needed. Electrolyte abnormalities may exist either during the acute stage or can develop readily in patients with low salt intake or during diuresis. Since thromboembolism from endocardial thrombi is a prominent feature, occurring in as many as 80% of individuals in one series (73), anticoagulants are important. Finally, where facilities are available, prolonged bed rest can reduce the size of an excessively dilated heart and, more importantly, may contribute to control of alcohol intake (74).

CONCLUSION

The widespread use of ethyl alcohol suggests its potential importance in clinical medicine. There is no proven therapeutic effect in cardiac patients and its role as an etiologic factor in heart disease has been disputed over the years and attributed to coexistent malnutrition. The latter factor, however, has been dissociated from ethanol use in many patients with the cardiomyopathic form of heart failure.

Major support for the role of ethanol as a toxic agent when used in large amounts for a prolonged period has been obtained in various species of animals, including the subhuman primate. Abnormalities include depression of ventricular function, and metabolic and morphologic changes that parallel the changes in humans with preclinical malfunction of the heart. While the mechanism of progression to heart failure or arrhythmias is not known, several factors may be associated. These include, particularly in males, the cumulative effects of ethanol alone or after intensified drinking episodes, simultaneous exposure to trace metals in excess, and occa-

Table 1 Course of alcoholic cardiomyopathy (3.3 years); $N = 57$

Group	Status	Deaths	Abstained from Alcohol
A	improved	0	11/15
B	same	0	3/12
C	worse	24	4/30

sional specific nutritional deficiency or superimposed infection. The low prevalence of clinical nutritional deficiency in patients with alcoholic cardiomyopathy and the infrequency of heart disease in patients with cirrhosis or neuropathy supports the view that the cardiac abnormality is commonly not dependent on malnutrition.

Clinical data indicate that the cessation of alcohol intake may reverse the disease or interrupt its progression in many patients. However, the pathogenetic process may continue unabated in some patients who become abstinent.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the secretarial services rendered by Audrey Brown and Anne Binetti.

Literature Cited

1. Ferrans, V. J., Rios, J. C., Gooch, A. S., Nutter, D., DeVita, V. T., Datlow, D. W. 1966. Alcoholic cardiomyopathy. *Am. J. Med.* 252:123–36
2. Brigden, W., Robinson, J. 1964. Alcoholic heart disease. *Br. Med. J.* 2:1283–89
3. Burch, G. E., Walsh, J. J. 1960. Cardiac insufficiency in chronic alcoholism. *Am. J. Cardiol.* 6:864–74
4. Fowler, N. O., Gueron, M., Rowlands, D. T. 1961. Primary myocardial disease. *Circulation* 23:498–508
5. Neville, J. N., Eagles, J. A., Sampson, G., Olson, R. E. 1968. The nutritional status of alcoholics. *Am. J. Clin. Nutr.* 21:1329–40
6. Leevy, C. M., Baker, H., Ten-Hove, O., Frank, O., Cherrick, G. R. 1965. B-complex vitamins in liver disease of the alcoholic. *Am. J. Clin. Nutr.* 16:339–46
7. Cahalan, D., Cisin, I. H., Crossley, H. M. 1969. American drinking practices: A national survey of drinking behavior and attitudes. New Brunswick, NJ: Rutgers Univ. Cent. Alcohol Stud. Monogr. No. 6
8. Regan, T. J., Levinson, G. E., Oldewurtel, H. A., Frank, M. J., Weisse, A. B., Moschos, C. B. 1969. Ventricular function in noncardiacs with alcoholic fatty liver: the role of ethanol in the production of cardiomyopathy. *J. Clin. Invest.* 48:397–407
9. Ahmed, S. S., Levinson, G. E., Regan, T. J. 1973. Depression of myocardial contractility with low doses of ethanol in normal man. *Circulation* 48:378–85
10. Symbas, P. N., Tyras, D. H., Ware, R. E., Baldwin, B. J. 1972. Alteration of cardiac function by hemodialysis during experimental alcohol intoxication. *Circulation* 45, 46:II-227 (Abstr.)
11. Regan, T. J. 1971. Ethyl alcohol and the heart. *Circulation* 44:957–63
12. Conway, N. 1968. Haemodynamic effects of ethyl alcohol on normal human volunteers. *Br. Heart J.* 30:638–44
13. Newman, W. H., Valicenti, J. E. Jr. 1971. Ventricular function following acute alcohol administration: A strain-gauge analysis of depressed ventricular dynamics. *Am. Heart J.* 81:61–68
14. Wong, M. 1973. Depression of cardiac performance by ethanol unmasked during autonomic blockade. *Am. Heart J.* 86:508–15
15. Musacchio, J., Kopin, I. J., Snyder, S. 1964. Effects of disulfiram on tissue norepinephrine content and subcellular distribution of dopamine, tyramine and their β-hydroxylated metabolites. *Life Sci.* 3:769–75
16. Nokono, J., Holloway, J. E., Schackford, J. S. 1969. Effects of disulfiram on the cardiovascular responses to ethanol in dogs and guinea pigs. *Toxicol. Appl. Pharmacol.* 14:439–46
17. Hirose, T., Higashi, R., Ikeda, H., Tamura, K., Suzuki, T. 1973. Effect of ethanol on adrenaline and noradrenaline secretion of the adrenal gland in the dog. *Tohoku J. Exp. Med.* 109:85–88
18. Seeman, P., Chau, M., Goldberg, M., Sauks, T., Sax, L. 1971. The binding of Ca^{2+} to cell membrane increased by volatile anesthetics (alcohols, acetone, ether) which induce sensitization of nerve or muscle. *Biochim. Biophys. Acta* 225:185–93
19. Kalant, H., Israel, Y. 1967. Effects of ethanol on active transport of cations.

In *Biochemical Factors in Alcohol*, ed. R. P. Mickel, pp. 25–37

20. Regan, T. J., Koroxenidis, G. T., Moschos, C. B., Oldewurtel, H. A., Lehan, P. H., Hellems, H. K. 1966. The acute metabolic and hemodynamic responses of the left ventricle to ethanol. *J. Clin. Invest.* 45:270–80
21. Ferrans, V. J., Hibbs, R. G., Weilbaecher, D. G., Black, W. C., Walsh, J. J., Burch, G. E. 1965. Alcoholic cardiomyopathy: A histochemical study. *Am. Heart J.* 69:748–65
22. Gimeno, M. A. F., Bedners, A. S., de Vastik, F. J., Gimeno, A. L. 1971. Effect of ethanol on the motility of isolated rat myometrium. *Arch. Int. Pharmacodyn. Ther.* 191:213–19
23. Cooper, S. A., Dretchen, K. L. 1975. Biphasic action of ethanol on contraction of skeletal muscle. *Eur. J. Pharmacol.* 31:232–36
24. Kiessling, H.- H., Pilstrom, L., Bylund, A.- C., Piehl, K., Saltin, B. 1975. Effects of chronic ethanol abuse on structure and enzyme activities of skeletal muscle in man. *Scand. J. Clin. Lab. Invest.* 35:601–7
25. Knochel, J. P., Bilbrey, G. L., Fuller, T. J., Carter, N. W. 1975. The muscle cell in chronic alcoholism: The possible role of phosphate depletion in alcoholic myopathy. *Ann. NY Acad. Sci.* 252:274–86
26. Schenk, E. A., Cohen, J. The heart in chronic alcoholism. *Pathol. Microbiol.* 35:96–104
27. Maines, J. E., Aldinger, E. E. 1967. Myocardial depression accompanying chronic consumption of alcohol. *Am. Heart J.* 73:55–63
28. Lochner, A., Cowley, R., Brink, A. J. 1969. Effect of ethanol on metabolism and function of perfused rat heart. *Am. Heart J.* 78:770–80
29. Regan, T. J., Khan, M. I., Ettinger, P. O., Haider, B., Lyons, M. M., Oldewurtel, H. A. 1974. Myocardial function and lipid metabolism in the chronic alcoholic animal. *J. Clin. Invest.* 54:740–52
30. Ettinger, P. O., Lyons, M., Oldewurtel, H. A., Regan, T. J. 1976. Cardiac conduction abnormalities produced by chronic alcoholism. *Am. Heart J.* 91:66–78
31. Pachinger, O. M., Tillmanns, H., Mao, J. C., Fauvel, J.- M., Bing, R. J. 1973. The effect of prolonged administration of ethanol on cardiac metabolism and performance in the dog. *J. Clin. Invest.* 52:2690–96
32. Joly, J.-G., Feinman, L., Ishii, H., Lieber, C. S. 1973. Effect of chronic ethanol feeding on hepatic microsomal glycerophosphate acyl transferase activity. *J. Lipid Res.* 14:337–43
33. Reitz, R. C., Helsabeck, E., Mason, D. P. 1973. Effects of chronic alcohol ingestion on the fatty acid composition of the heart. *Lipids* 8:80–84
34. Hall, J. L., Rowlands, D. T. Jr. 1970. Cardiotoxicity of alcohol. An electron microscopic study in the rat. *Am. J. Pathol.* 60:153–60
35. Gvozdjak, A., Bada, V., Kruty, F., Niederland, T. R., Gvozdjak, J. 1973. Chronic effect of ethanol on the metabolism of myocardium. *Biochem. Pharmacol.* 22:1807–11
36. Bulloch, R. T., Pearce, M. B., Murphy, M. L., Jenkins, B. J., Davis, J. L. 1972. Myocardial lesions in idiopathic and alcohol cardiomyopathy. Study by ventricular septal biopsy. *Am. J. Cardiol.* 29:15–25
37. Ferrans, V. J., Roberts, W. C., Shugoll, G. I., Massumi, R. A., Ali, N. 1973. Plasma membrane extensions in intercalated discs of human myocardium and their relationship to partial dissociations of the discs. *J. Mol. Cell Cardiol.* 5:161–69
38. Hibbs, R. G., Ferrans, V. J., Black, W. C., Weilbaecher, D. G., Walsh, J. J., Burch, G. E. 1965. Alcoholic cardiomyopathy. An electron microscope study. *Am. Heart J.* 69:766–79
39. Alexander, C. S. 1966. Idiopathic heart disease. II. Electron microscopic examination of myocardial biopsy specimens in alcoholic heart disease. *Am. J. Med.* 41:229–34
40. Auger, C., Chenard, J. 1967. Quebec beer-drinkers' cardiomyopathy: ultrastructural changes in one case. *Can. Med. Assoc. J.* 97:916–21
41. Vasdev, S. C., Chakravarti, R. N., Subrahmanyam, D., Jain, A. C., Wahi, P. L. 1975. Myocardial lesions induced by prolonged alcohol feeding in rhesus monkeys. *Cardiovasc. Res.* 9:134–40
42. Sagel, L. D., Rendig, S. V., Choquet, Y., Chacko, K., Amsterdam, E. A., Mason, D. T. 1975. Effects of chronic graded ethanol consumption on the metabolism, ultrastructure, and mechanical function of the rat heart. *Cardiovasc. Res.* 9:649–63
43. Burch, G. E., Harb, J. M., Colcolough, H. L., Tsui, C. Y. 1971. The effect of

prolonged consumption of beer, wine and ethanol on the myocardium of the mouse. *Johns Hopkins Med. J.* 129:130–48

44. Jesrani, M. U., Gopinathan, K., Khan, M. I., Oldewurtel, H. A., Regan, T. J. 1971. Acetaldehyde and the myocardial depressant effects of ethanol. *Circulation* 44:II–127 (Abstr.)
45. Jones, J. E., Shane, S. R., Jacobs, W. H., Flink, E. B. 1969. Magnesium balance in chronic alcoholism. *Ann. NY Acad. Sci.* 162: Suppl. 2, pp. 934–46
46. Limas, C. J., Guiha, N. H., Lekagul, O., Cohn, J. N. 1974. Impaired left ventricular function in alcoholic cirrhosis with ascites. Ineffectiveness of ouabain. *Circulation* 49:755–60
47. Gould, L., Shariff, M., DiLieto, M. 1969. Cardiac hemodynamics in alcoholic patients with chronic liver disease and presystolic gallop. *J. Clin. Invest.* 48:860–68
48. Spodick, D. H., Pigott, V. M., Chirife, R. 1972. Preclinical cardiac malfunction in chronic alcoholism. Comparison with matched normal controls and with alcoholic cardiomyopathy. *N. Engl. J. Med.* 287:677–80
49. Wu, C. F., Sudhakar, M., Jaferi, G., Ahmed, S. S., Regan, T. J. 1976. Preclinical cardiomyopathy in chronic alcoholics: A sex difference. *Am. Heart J.* 91:281–86
50. Regan, T. J., Wu, C. F., Weisse, A. B., Moschos, C. B., Ahmed, S. S., Lyons, M. M., Haider, B. 1975. Acute myocardial infarction in toxic cardiomyopathy without coronary obstruction. *Circulation* 51:453–61
51. Hognestad, J., Teisberg, P. 1973. Heart pathology in chronic alcoholism. *Acta Pathol. Microbiol. Scand.* 81(3):315–22
52. Galambos, J. T. 1974. Alcoholic hepatitis. In *The Liver and Its Diseases,* ed. F. Schaffner, S. Sherlock, C. M. Leevy, pp. 255–67. New York: Intercont. Med. Book Corp.
53. Pequignot, G. 1961. Die rolle des alkohols bei der antiologies von Leberzirrhosen in Frankreich. *Muench. Med. Wochenschr.* 103:1464–68
54. Albot, G., Boisson, J., Chaput, J. C., Riviere, M. 1968. Etude comparative des facteurs toxiques et nutritionnels dans les hepatities alcooliques subaugues. *Rev. Int. Hepatol.* 18:89–104
55. Martini, G. A., Bode, C. 1970. The epidemiology of cirrhosis of the liver. In *Alcoholic Cirrhosis and Other Toxic Hepatopathies,* ed. A. Engle, T. Larsson, pp. 315–35. Skand. Int. Symp. Stockholm: Nord. Bokhandelns Forlag
56. Ugarte, G., Iturriaga, H., Insunza, I. 1970. Some effect of ethanol on normal and pathologic livers. In *Progress in Liver Disease,* ed. H. Popper, F. Schaffner, 3:335–70. New York: Grune & Stratton
57. Caroli, J., Pequignot, G. 1958. Enquete sur les circonstances dietetiques de la cirrhose alcoolique en France. *Proc. World Congr. Gastroenterol.* 1:661
58. Mackay, I. R., Langford, L. 1963. Blood alcohol estimation in patients attending an outpatient clinic. *Med. J. Aust.* 1:778–79
59. Lelback, W. K. 1966. Leberschaden bei chronischem alkoholismus. *Acta Hepato-Splenol.* 13:321–49
60. Sherlock, S., Brunt, P., Scheurer, P. J. 1970. Clinical and pathological aspects of alcoholic liver disease. In *Alcohol and the Liver,* ed. W. Gerok, K. Sickinger, H. H. Hennekeuser, pp. 383–88. Stuttgart: Schattauer
61. Wenger, N. J., Bauer, S. 1958. Coronary embolism. *Am. J. Med.* 25:549–57
62. Cleland, J., Molloy, P. J. 1973. Thromboembolic complications of the cloth-covered Starr-Edwards prosthesis No. 2300 aortic and No. 6300 mitral. *Thorax* 28:41–47
63. Harker, L. A., Slichter, S. J. 1972. Platelet and fibrinogen consumption in man. *N. Engl. J. Med.* 287:999–1005
64. Moschos, C. B., Lahiri, K., Manskopf, G., Oldewurtel, H. A., Regan, T. J. 1973. Effect of experimental coronary thrombosis upon platelet kinetics. *Thromb. Diath. Haemorrh.* 30:339–46
65. Ettinger, P. O., Wu, C. F., de La Cruz, C. Jr., Weisse, A. B., Regan, T. J. 1976. Tachyarrhythmias associated with the preclinical cardiomyopathy of alcoholism. *Am. J. Cardiol.* 37:134 (Abstr.)
66. Evans, W. 1959. The electrocardiogram of alcoholic cardiomyopathy. *Br. Heart J.* 21:445–56
67. Levine, H. D., Piemme, T. E., Monroe, K. E. 1965. A brisk electrocardiogram observed in chronic alcoholics. *Am. Heart J.* 69:140–42
68. Horan, L. G., Flowers, N. C., Thomas, J. R., Tolleson, W. J. 1964. The spatial vectorcardiogram in idiopathic cardiomyopathy. *Prog. Cardiovasc. Dis.* 7:115–24
69. Kramer, K., Kuller, L., Fisher, R. 1968. The increasing mortality attributed to cirrhosis and fatty liver, in

Baltimore (1957–1966). *Ann. Intern. Med.* 69:273–82

70. Sundby, P. 1967. Alcoholism and mortality. In *Universitetsforlaget* (Alcohol research in the northern countries), Publ. 6, Stockholm: National Institute for Alcohol Research; New Brunswick, NJ: Rutgers Univ. Cent. Alcohol Stud.
71. De La Cruz, C. L. Jr., Haider, B., Ettinger, P. O., Regan, T. J. 1976. Effects of ethanol on ventricular electrical stability in the chronic alcoholic animal. *Clin. Res.* 24(3):419A (Abstr.)
72. Demakis, J. G., Proskey, A., Rahimtoola, S. H., Jamil, M., Sutton, G. C., Rosen, K. M., Gunnary, R. M., Tobin, J. R. 1974. The natural course of alcoholic cardiomyopathy. *Ann. Intern. Med.* 80(3):293–97
73. Morin, Y. Q. 1967. Beer-drinkers' cardiomyopathy: Hemodynamic alterations. *Can. Med. Assoc. J.* 97:901–4
74. Burch, G. E., DePasquale, N. P. 1969. Alcoholic cardiomyopathy. *Am. J. Cardiol.* 23:723–31

Ann. Rev. Med. 1977. 28:411-23
Copyright © 1977 by Annual Reviews Inc. All rights reserved

THE CURRENT STATUS OF BCG IMMUNIZATION AGAINST TUBERCULOSIS

❖7252

Theodore C. Eickhoff, M.D.
Department of Medicine, University of Colorado Medical Center, Denver, Colorado 80262

INTRODUCTION

The use of BCG vaccine in the prevention of tuberculosis has long been a controversial and emotionally charged issue in the United States. Although available in this country, BCG vaccine has been used but little in the prevention of tuberculosis. In contrast, BCG vaccination has been the major thrust of the World Health Organization's efforts to control tuberculosis in countries with high rates of transmission, and is credited with a major role in reducing tuberculosis morbidity.

In controlling tuberculosis in the United States, reliance has been placed by public health authorities on early detection of new cases using the tuberculin skin test, prompt chemotherapy, and preventive isoniazid treatment of household contacts, skin-test converters, and other high-risk groups with positive skin tests.

The validity of this approach has been evident, for tuberculosis has declined sharply in the United States during the past several decades. United States Public Health Service data (1, 2) indicate that in 1953 there were 84,000 new cases of tuberculosis and 19,700 deaths due to tuberculosis; in 1975 there were only 33,500 new cases, and the number of tuberculosis deaths had declined to 3,500. Other, more general factors, such as improved nutrition and the gradual increase in standard of living that has characterized the United States economy, have unquestionably contributed to the decline in tuberculosis morbidity. There remain, however, localized foci or "pockets" of tuberculosis transmission in the United States, particularly in areas in which preventive medical services are less well developed or cannot be effectively delivered.

This review examines the history of BCG vaccine, the evidence that can be brought to bear on its efficacy and safety, and the current problems and unanswered questions that attend its use, and finally suggests circumstances under which BCG vaccine may be appropriately used in the United States.

The use of BCG vaccine as a nonspecific immunostimulant in cancer chemotherapy is the subject of a rapidly increasing body of literature. This review focuses only on the use of BCG in preventing tuberculosis; there are, however, some problems common to both uses of this product.

RATIONALE FOR VACCINATION AGAINST TUBERCULOSIS

Earlier in this century, most people eventually became infected with tubercle bacilli, as demonstrated by skin-test positivity. However, only a small proportion of those who were infected developed overt tuberculous disease. Most people who were infected appeared to have acquired a degree of resistance against developing tuberculosis upon subsequent exposure, which earlier in this century was frequent and virtually unavoidable.

As a result of the immunologic studies of Mackaness (3–7) and the clinical work of Stead (8–13), among others, immunity in tuberculosis is now better understood in terms of modern immunologic concepts, and the "unitary concept" of the pathogenesis of tuberculosis in man is generally accepted. Thus, primary infection with tubercle bacilli results in activation of host cell-mediated immune mechanisms, and is reflected clinically in the appearance of a positive tuberculin skin test. If the primarily infected person has received a large dose of tubercle bacilli, or if cell-mediated immune mechanisms do not respond optimally, the individual may go on to develop overt clinical tuberculosis. Most frequently, however, the tuberculous infection is localized by the host cell-mediated immune mechanisms, resulting in a dormant or latent infection, which may (*a*) remain dormant for life, or (*b*) reactivate at some time in the future. Reactivation is frequently, but not invariably, associated with conditions known to modify host cell-mediated immune mechanisms, such as malnutrition, diabetes, malignancy, or steroid therapy.

There is abundant clinical and experimental evidence that tuberculin positivity, reflecting activated cell-mediated immune mechanisms, confers protection against exogenous exposure to tuberculosis (8). Such individuals are, however, at risk of reactivation or "breakdown" tuberculosis. Tuberculin-negative individuals are susceptible to primary infection, but by definition are not at risk of reactivation tuberculosis.

The use of BCG vaccine, an attenuated strain immunologically closely related to virulent *Mycobacterium tuberculosis*, attempts to gain the advantage of protection conferred by activated host cell-mediated immune mechanisms without risking progressive disease in man.

HISTORY OF BCG VACCINE

The bacillus of Calmette and Guérin, known as BCG, was originally derived from a virulent strain of *Mycobacterium bovis*, attenuated by 231 serial passages over a period of 13 years on beef bile–containing medium (14). The early studies of Calmette & Guérin at the Pasteur Institute showed that animals immunized with this culture developed increased resistance to a challenge dose of virulent tubercle bacilli (14). BCG vaccine was first administered by mouth to newborn infants in 1921 (15), and since then has been administered to more than a 100 million persons of all ages (16).

The organism was maintained by serial passage at the Pasteur Institute, and in the decades following its description, was subcultured and distributed to hundreds of laboratories in many countries. In those laboratories, many of which produced their own BCG vaccines, the strain was similarly maintained by serial subculture. It became apparent over 25 years ago that serial subculturing on differing media in many different laboratories had resulted in the production, by inadvertent selection, of many different "daughter" BCG strains that differed, sometimes widely, in gross morphology, growth characteristics, biochemical activity, sensitizing potency, and even animal virulence (17–23). It was not possible, of course, to carry out direct comparisons of any of the BCG "daughter" strains to the original bacillus of Calmette & Guérin. In the last two decades, most production laboratories have adopted a seed-lot system, maintaining production strains in a lyophilized state, in an attempt to minimize the genetic variation that is unavoidable in serial subculture (24). The situation currently is thus one of many laboratories producing BCG vaccine using their own "daughter" strain, preserved in a seed-lot system. The production strains are generally named by the city in which the production laboratory is located, e.g. Paris, Copenhagen, London, Montreal, Rio de Janeiro, etc. Thus there is no one BCG vaccine; there are, rather, dozens of different BCG "daughter" vaccines.

VACCINE PRODUCTION

Current requirements of the Bureau of Biologics, US Food and Drug Administration, specify that the vaccine production strain must have been maintained as a lyophilized primary seed lot, but do not specify the use of any particular strain. It is required only that the source of the vaccine strain be identified by complete historical records (25).

In most production laboratories, the bacilli are grown as a pellicle on the surface of liquid Sauton medium, or dispersed throughout Sauton medium. An early harvest, 6–9 days, is considered important for good survival after freeze-drying. After filtering and pressing, the semi-dry mycobacterial mass is homogenized at a controlled temperature, diluted, and subsequently freeze-dried (26).

Routine quality control carried out by production laboratories generally includes an identity test; tests for contamination; safety test in guinea pigs; estimate of total bacillary mass by opacity and dry weight; viability determined by oxygen uptake, germination rate, or colony count; and tests of heat stability. Such routine tests are particularly important for insuring lot-to-lot uniformity.

Additional proposed federal requirements, not yet formally adopted, include the necessity of demonstrating that production lots of BCG vaccine are incapable of producing progressive tuberculosis in guinea pigs, and induce tuberculin skin-test positivity using 5–10 tuberculin units (TU) of purified protein derivative (PPD) in

90% or more of a sample of persons, previously tuberculin-negative, after being given BCG vaccine (25).

SAFETY OF BCG VACCINE

The early history of BCG vaccination was tarnished by the Lubeck catastrophe, in which 72 of 251 infants died of tuberculosis following BCG vaccination. That disastrous episode was subsequently shown to be due to contamination of the vaccine by a strain of virulent tubercle bacilli (27). Excluding that episode, therefore, the safety of BCG vaccine has never been a serious issue. Progressive disease has occasionally been reported in immunosuppressed hosts, particularly in hosts with defects of cell-mediated immune mechanisms (28). In summarizing the world's literature through 1968, however, Mande could find only 13 fatalities reported as being due to BCG vaccination (29). In countries in which BCG is regularly used in the immunization of infants, for example Sweden and Finland, osteomyelitis due to BCG has recently been reported with a frequency as high as 5 per 100,000 among newborns (30). Although worrisome in areas where BCG is regularly used to vaccinate infants, this finding does not substantially detract from the generally excellent safety record of BCG vaccine.

EFFICACY OF BCG VACCINES IN HUMANS

Sutherland notes seven large, acceptably controlled trials of BCG vaccine (31). These are summarized in Table 1. Since these are the critical trials, they are described in some detail, as are a few additional trials that provide relevant information.

Field Trials Showing Excellent Protection

The first was reported by Stein & Aronson in 1953 (32), and carried out among eight North American Indian tribes. The period of intake into the study was 1935 to 1938, and individuals from 0 to 20 years were included. The criterion of eligibility was negativity to 250 T.U. of PPD. The source of the vaccine was the Henry Phipps Institute in Philadelphia. The duration of follow-up was 9 to 11 years; there were approximately 1500 subjects each in the unvaccinated and BCG groups. Two hundred and thirty-eight cases of tuberculosis occurred in the unvaccinated group, and 64 in the BCG group; the protective efficacy was calculated to be approximately 80%.

Another study showing a high degree of protection was that carried out by Rosenthal et al, reported in 1961 (33), of Chicago infants in high-risk areas. The period of intake was 1937 to 1948. Children under 3 months were immunized with no initial tuberculin testing. The source of the vaccine was the Tice Laboratory in Chicago, and follow-up ranged from 12 to 23 years. There were approximately 1700 subjects each in the unvaccinated and BCG groups. Sixty-five cases of tuberculosis occurred among the unvaccinated, and 17 among the BCG group. The protective efficacy was found to be 75%.

Table 1 Summary of seven large controlled trials of BCG immunization against tuberculosis[a]

Trial	Reference	Intake period	Vaccine laboratory	Duration of observation (years)	Percentage protection from BCG
North America: Indians	Stein & Aronson 1953 (32)	1935–38	Phipps	9-11	80
Chicago: infants	Rosenthal et al 1961 (33)	1937–48	Tice	12–23	75
Britain: schoolchildren	Medical Research Council 1972 (34)	1950–52	Copenhagen	15	78
Georgia: schoolchildren	Comstock & Webster 1969 (38)	1947	Tice	20	–
Puerto Rico: children	Palmer et al 1958 (35)	1949–51	NY State	5½–7½	31
Georgia, Alabama: population	Comstock & Palmer 1966 (39)	1950	Tice	14	14
Southern India: rural population	Frimodt-Møller et al 1964 (37)	1950–55	Madras	2½–7	60/31

[a] Adapted from Sutherland (31).

The third trial showing excellent protection was that from the Medical Research Council, (MRC) of Great Britain, as reported most recently in the WHO Bulletin in 1972 (34). The period of intake was from 1950 to 1952, and the subjects were 14- to 15-year-old British school-leavers. The criterion for eligibility was a skin-test response less than 5 mm to 100 T.U. of old tuberculin. The source of vaccine was the Statens Seruminstitut in Copenhagen, and the duration of follow-up was 15 years. There were approximately 13,000 subjects each in the unvaccinated and BCG groups. Two hundred and forty cases of tuberculosis occurred among the unvaccinated, and 56 in the BCG-vaccinated; the protective efficacy was 78%.

Field Trials Showing Mediocre Protection

Two studies show wholly mediocre protection. The first was the study reported by Palmer et al, in 1958 (35, 36), from the general population in Puerto Rico, one of the USPHS-sponsored BCG trials. The period of intake was from 1949 to 1951, and included children from 1 to 18 years. The criterion for eligibility was a reaction less than 6 mm to 10 T.U. of PPD. The vaccine was prepared by the State Department of Health in Albany, New York. The duration of follow-up varied from 5½ to 7½ years, with a mean of 6.3 years. There were approximately 27,000 in the unvaccinated population and 51,000 in the BCG-vaccinated population. Seventy-three cases of tuberculosis occurred in the unvaccinated group, and 93 in the BCG group, a protective efficacy of 31%.

The second study showing mediocre protection was reported by Frimodt-Møller et al, in 1964 (37), carried out in a rural population in South India. The intake period was 1950–1955, and included all ages. The criterion for eligibility was a reaction of less than 5 mm to 5 T.U. of old tuberculin. The source of the vaccine was the BCG laboratory in Madras, India, and the duration of follow-up was 2½–7 years. There were approximately 6000 people in the unvaccinated group, and 5000 in the BCG group. The initial estimate of efficacy was 60%. In a subsequent report, however, cited in the MRC study (34), when follow-up was extended to 9–14 years, the efficacy figure declined to 31%.

Field Trials Showing Poor or No Protection

Two studies cited by Sutherland show little or no effect of BCG. These were both USPHS-sponsored studies.

The first is the Muscogee County Study in Georgia, as reported by Comstock & Webster in 1969 (38). The period of intake was 1947, and included children of age 6–17 years. The criterion for eligibility was a reaction of less than 5 mm to 100 T.U. of PPD. The source of the vaccine was the Tice Laboratory in Chicago, and the duration of follow-up was 20 years. There were approximately 2300 subjects in the unvaccinated group and 2500 in the BCG-vaccinated group. Three cases of tuberculosis occurred in the unvaccinated group, and 5 in the BCG group, yielding a protective efficacy of nil.

The second study was carried out in the general population of Georgia and Alabama, reported by Comstock & Palmer in 1966 (39). The period of intake was 1950, and included individuals over 5 years of age. The criterion for eligibility was a reaction of less than 5 mm to 5 T.U. of PPD tuberculin. The vaccine was again provided by the Tice Laboratory in Chicago, and the duration of the follow-up was 14 years. There were approximately 18,000 individuals in the unvaccinated group, and 17,000 in the BCG group. Thirty-two cases of tuberculosis were observed in the unvaccinated group and 26 in BCG group, yielding a protective efficacy of 14%.

Thus, in the seven controlled field trials cited, a strikingly wide range of protective efficacy was seen, ranging from 0 to 80%. As far as it is possible to judge, each of these trials was methodologically reliable, and provided a valid estimate of the efficacy of the trial vaccine in the population in which it was used. They have not, to date, been successfully challenged on scientific grounds, in spite of a recent unconvincing editorial attempt to challenge the MRC trial on the grounds of bias in follow-up (40). Methods of case detection have been variable, however, and are critically important in those trials in which the detected incidence of tuberculosis in the control group was already quite low.

Several additional smaller-scale controlled or uncontrolled studies provide relevant, albeit contrasting, information.

In a study of Canadian Indian infants vaccinated at birth during the years 1933–1945 (41), BCG conferred protection in a highly infectious environment, 68.7% of children being infected at an average age of 9. The vaccine used was prepared by Dr. Armand Frappier, Director of the Institute of Microbiology and Hygiene at the University of Montreal. There were 306 vaccinated infants and 303 controls, randomly allocated. The percentage of manifest tuberculosis was in the ratio of 1:4.85 comparing the vaccinated with controls, an efficacy of almost 80%.

Quite in contrast, Levine & Sackett reported in 1946 (42) the results of their studies of BCG vaccination of infants in New York City. The study ranged over the period 1926 to 1944, and the vaccine used was prepared by the Phipps Institute in Philadelphia, the same source of vaccine as used in the Aronson study in American Indians. Between 1926 and 1932, the infants were divided into two groups according to whether the mother wished her baby to be vaccinated or not. Tuberculosis mortality was reduced almost 6-fold in the vaccinated group as compared to the controls. During the period from 1933 to 1944, the mode of allocation into groups was randomized as follows. Consecutive mothers were asked whether they would like to have their children vaccinated. Those who refused were not considered further. To the children of the mothers who were willing to accept vaccination, BCG vaccination was given alternately without selection. Under these circumstances, with over 500 children in each group, the vaccine efficacy appeared to be nil.

Bettag and associates reported in 1964 (43) the results of a BCG study done at the Lincoln State School, a state institution for the mentally retarded in Illinois. Vaccination was carried out in 1947 among participants determined to be negative to both 1:1000 and 1:100 old tuberculin. The BCG used was prepared and administered by the Tice Laboratory in Chicago. A follow-up evaluation was made 13 years later, in 1960. The evaluation revealed that 12 of the BCG-vaccinated individuals had developed tuberculosis, 5 of whom died. Among the control group of approximately equal size, 8 had developed tuberculosis, with 3 deaths. There were thus more cases of tuberculosis among the vaccinated than among the nonvaccinated, although the deviation from zero protection was not significant.

Again in contrast, Hyge reported in 1957 (44) an unusual tuberculosis epidemic observed in a Danish state school for young girls. The epidemic resembled a controlled experiment in that it involved 105 initially tuberculin-negative girls, 133 BCG-vaccinated girls, and 130 initially tuberculin-positive girls, nearly all of whom were exposed to the same massive infection under the same conditions. Of the 105 tuberculin-negative girls, 70 converted to positive, and almost 60% of the converters developed primary tuberculosis. Almost one third of those with primary tuberculosis developed chronic pulmonary tuberculosis. Of the 133 BCG-vaccinated girls, there were no cases of primary tuberculosis, and only two girls who subsequently developed chronic pulmonary tuberculosis. Among the 130 originally tuberculin-positive girls, 9 cases of progressive pulmonary tuberculosis appeared during the 12 years of observation. There were thus eight times as many cases of post-primary tuberculosis in the tuberculin-negative group as in the BCG-vaccinated group, and the total incidence of tuberculosis, primary plus post-primary, was 23 times as high among the tuberculin-negative as among the BCG-vaccinated girls. The source of the vaccine used was not specifically stated, but was most probably from the Statens Seruminstitut in Copenhagen.

How can such widely disparate results be explained? Three possibilities seem to merit serious attention.

First is the explanation for the poor results found in the Georgia-Alabama trials by Palmer and his colleagues. Palmer suggested (45) that in areas where nonspecific tuberculin sensitivity was common, as is true throughout much of the southeastern

United States, a large proportion of the population had already acquired some natural immunity against virulent tuberculous infection from atypical mycobacterial infections. In this situation, vaccination with BCG would only supplement the immunity that already existed and would make a relatively smaller apparent contribution to protection than in areas that were essentially free from atypical mycobacterial infections. This hypothesis has been experimentally supported in guinea pigs, showing that infection with other mycobacteria did indeed confer protection against subsequent virulent challenge (45). This protection, however, was always less than was conferred by BCG. Palmer suggested that this explanation could, at least in part, reconcile the widely differing findings of the MRC trial in Great Britain and those in the Southeastern United States.

Hart, however, subsequently showed that while differences in the frequency of other mycobacterial infections could well have contributed to this difference, it was unlikely to be the whole story (46). If none of the subjects in the Georgia-Alabama trial had any natural protection from other mycobacterial infections, the apparent efficacy of the vaccine in that population would have risen from the actual 14% to only 25%. He postulated that some other influence must be operating, and suggested that the vaccine used in the Georgia-Alabama trial must have been less potent than the Danish strain used in the MRC trial.

This is the second possibility that merits attention, namely that different products all labeled as BCG may differ widely in their immunizing effect, and that this could be the main reason, or even the only one, for the mutally contradictory results of different BCG trials.

At this date, it is difficult, if not impossible, to ascertain whether the vaccines, the technique of administration, or both were responsible for the divergent results noted in controlled field trials. There is corroborating evidence, however, that BCG strains used in vaccine production by the laboratory supplying vaccine for two of the field trials that showed no protection were very weak in terms of multiplication, allergenic potency, and protection in animals (47, 48).

The third possibility is one recently suggested by Sutherland (31). He had observed that areas with a high incidence of tuberculosis in the unvaccinated group showed a high efficacy of BCG vaccine, whereas those with a low incidence of tuberculosis in the unvaccinated group showed a low efficacy, suggesting that the efficacy of BCG may be greater in an area where there is much tuberculosis than in an area where there is only little. If this relationship is genuine, it suggests that the superinfection of vaccinated subjects with virulent tubercle bacilli or other mycobacteria may be necessary to maintain the protection conferred by BCG vaccine. This concept is not without its parallels in other infectious diseases, but has not heretofore been suggested for tuberculosis and BCG vaccine. There is no direct evidence by which to judge this hypothesis, and the suggested relationship between endemic infection rate and protection from BCG must remain no more than an intriguing possibility for the present.

All of the controlled field trials cited previously were carried out using liquid BCG vaccines. There have thus far been no controlled field trials of freeze-dried BCG vaccines reported, though one is currently in progress in India. To date the only evidence supporting the efficacy in man of freeze-dried BCG vaccine is wholly uncontrolled. The results of that trial suggest, but do not prove, that the freeze-dried vaccine prepared by Glaxo Laboratories is as effective in man as the liquid Copenhagen vaccine used in the MRC trial in Great Britain (49).

On the basis of presently available information, therefore, judgements concerning the efficacy of currently available BCG vaccines can be made only by inference.

CURRENT PROBLEMS AND UNANSWERED QUESTIONS

Marked differences in the immunogenic and sensitizing potency of BCG strains were demonstrated over 20 years ago (17–23). During continuous serial subculturing, the traditional way of maintaining strains prior to the introduction of seed-lot systems, the emergence of mutant strains was unavoidable. Mutants that have a faster growth rate in vitro than the parent cells can, in a relatively shorter period of time, emerge as the dominant strain. There have been striking spontaneous changes in such attributes as morphology, pigmentation, rate of growth, and even in the ability to protect animals against experimental infection. In the case of such marked phenotypic change, the "daughter" strain can no longer be regarded as the same as the parent strain. Seed-lot systems have been used to preserve BCG strains for little more than a decade. There is no single scientifically defined entity known as BCG vaccine; there are rather many different BCG vaccines, with varied biological characteristics, varied animal virulence, and almost surely varied immunizing potency in man (50–54).

The need for further strengthening of animal-model systems was highlighted by the recent report of Wiegeshaus and associates (55). In order to determine whether the method by which a vaccine was tested was a major factor contributing to the results, an experiment was conducted in which 5 different vaccines were distributed to each of 9 participating laboratories. Each investigator evaluated the potency of the vaccines in one or more animal models of his own choosing. Thus the method of vaccine preparation remained constant, but all other variables were permitted to change. The ranking of the 5 vaccines was essentially random, demonstrating that the method by which the vaccine is tested in animals markedly influences its apparent potency.

Nevertheless, many authorities consider that there is some correlation between the potency of vaccine for animals and its protective potency for man (54, 56). BCG vaccine with a high potency in animals may be expected to induce strong and long-lasting protection against tuberculosis in man, whereas a vaccine with low potency for animals may be virtually worthless for vaccination of humans. Thus it would seem reasonable to choose for the production of vaccines only strains that are metabolically fully active, have good immunogenic potency in animals, and induce strong and lasting tuberculin sensitivity in humans.

One further controlled field trial of BCG vaccine is currently in progress in India, supported by the World Health Organization and the US Public Health Service. This is the only controlled field trial of freeze-dried vaccines, and has utilized vaccines from several production laboratories at several dosage levels. This may well be the last opportunity to carry out well-controlled field trials of tuberculosis immunoprophylaxis, and the results will be awaited with considerable interest. Perhaps this trial will provide a scientific basis on which to recommend that all BCG vaccines be prepared from the same seed-lot strain of demonstrated efficacy in man.

A ROLE FOR BCG VACCINE TODAY?

This has long been a controversial issue in the United States. The proper use of BCG vaccine is to prevent tuberculosis, but controversy has arisen when attempts were made to define the groups of individuals or populations that would benefit from BCG vaccination.

Springett (57) has critically reviewed the circumstances under which the use of BCG vaccine should be considered. He concluded that in communities or age groups in which annual tuberculin conversion rates are 0.5–1% or more, BCG vaccination is justified on a cost-benefit basis. Where skin-test conversion rates are less than 0.5% annually, as is true in most of the United States, the tuberculosis attack rate in tuberculin-negative individuals becomes so low that BCG vaccination is no longer economically justified.

Since the present approach to tuberculosis control in the United States has in general been quite successful, and since transmission of tuberculosis in this country is already at a very low level, it is clear that BCG vaccine has no broad public-health role to play in preventing tuberculosis. The greatest risk of overt, active tuberculosis in this country resides in the tuberculin-positive population, and BCG vaccination would not alter this risk. Thus there is no rationale for mass use of BCG vaccine in this country.

There are, however, residual foci of tuberculosis transmission at high rates, and other selected instances of unavoidable repeated exposure of tuberculin-negative individuals to infectious patients. It is this kind of circumstance in which the use of BCG vaccine merits serious consideration. In contrast to previous highly restrictive statements (58), recent recommendations of advisory bodies such as the Public Health Service Advisory Committee on Immunization Practices (59) have adopted a rather more permissive stance, and have viewed BCG vaccine as an adjunct to tuberculosis control, supplementing methods of case detection, chemotherapy, and preventive treatment in those limited segments of the population in which an excessive rate of new infections can be demonstrated, and the usual surveillance and treatment programs have failed or cannot be readily applied.

The Committee on Infections Within Hospitals of the American Hospital Association has also adopted a policy of encouraging serious consideration of the use of BCG in health-care personnel in institutions in which the frequency of skin-test conversion exceeds 1% per year (60).

Given the present uncertainties, it would be reasonable to recommend the use of BCG vaccination under the following circumstances:

1. Infants who are unavoidably exposed to mothers or other close family contacts with active pulmonary tuberculosis.

2. Populations or defined groups in which there is evidence of high frequency of tuberculosis transmission, and in which the usual methods of case detection and isoniazid prophylaxis cannot be adequately carried out. Examples might include ghetto populations, alcoholics, migrants, and others without a regular source of health care.
3. Health-care personnel working in institutions serving major urban population centers, particularly those with large ghetto, skid-row, or derelict populations in which the endemic incidence of tuberculosis is relatively high, and especially if the skin-test conversion rate exceeds 1% annually.
4. Any tuberculin-negative individual at high risk of repeated exposure to infective cases.

Literature Cited

1. 1976. Morbidity and mortality weekly report: Annual supplement; summary 1975. Atlanta: Cent. Dis. Control
2. 1974. Reported tuberculosis data 1973. Atlanta: Cent. Dis. Control
3. Mackaness, G. B. 1968. The immunology of antituberculous immunity. *Am. Rev. Respir. Dis.* 97:337–44
4. Mackaness, G. B. 1971. Resistance to intracellular infection. *J. Infect. Dis.* 123:439–45
5. Mackaness, G. B. 1969. The influence of immunologically committed lymphocytes on macrophage activity *in vivo*. *J. Exp. Med.* 129:973–92
6. Collins, F. M., Mackaness, G. B. 1970. The relationship of delayed hypersensitivity to acquired antituberculous immunity. 1. Tuberculin sensitivity and resistance to infection in BCG-vaccinated mice. *Cell. Immunol.* 1:253–65
7. Mackaness, G. B. 1971. Delayed hypersensitivity and its significance. In *Status of Immunization in Tuberculosis in 1971*, ed. E. C. Chamberlayne, pp. 68–69. *Fogarty Int. Cent. Proc. No. 14.* DHEW Publ. No. (NIH) 72-68. Washington DC:GPO
8. Stead, W. W. 1965. The pathogenesis of pulmonary tuberculosis among older persons. *Am. Rev. Respir. Dis.* 91: 811–22
9. Stead, W. W. 1967. Pathogenesis of a first episode of chronic pulmonary tuberculosis in man: Recrudescence of residuals of the primary infection or exogenous reinfection? *Am. Rev. Respir. Dis.* 95:729–45
10. Stead, W. W. 1967. Pathogenesis of the sporadic case of tuberculosis (special article). *N. Engl. J. Med.* 277:1008–12
11. Stead, W. W., Kerby, G. R., Schlueter, D. P., Jordahl, C. W. 1968. Clinical spectrum of primary tuberculosis in adults: Confusion with reinfection in the pathogenesis of chronic tuberculosis. *Ann. Intern. Med.* 68:731–45
12. Stead, W. W., Bates, J. H. 1971. Evidence of a silent bacillemia in primary tuberculosis in man. *Ann. Intern. Med.* 74:559–61
13. Stead, W. W. 1971. The natural history of tuberculosis in man. See Ref. 7, pp. 13–18
14. Guerin, C. 1957. Early history of BCG. In *BCG Vaccination Against Tuberculosis*, ed. S. R. Rosenthal, pp. 48–53. Boston: Little, Brown
15. Weill-Halle, B. 1957. Oral vaccination. See Ref. 14, pp. 175–82
16. Crispen, R. G. 1974. BCG vaccine in perspective. *Semin. Oncol.* 1:311–17
17. Jensen, K. A. 1946. Practice of the Calmette vaccination. *Acta Tuberc. Scand.* 20:1–45
18. Jacox, R. F., Meade, G. M. 1949. Variation in the duration of tuberculin skin sensitivity produced by 2 strains of BCG. *Am. Rev. Tuberc.* 60:541–46
19. Dubos, R. J., Pierce, C. H., Schaefer, W. B. 1953. Antituberculous immunity induced in mice by vaccination with living cultures of attenuated tubercle bacilli. *J. Exp. Med.* 97:207–20
20. Dubos, R. J., Pierce, C. H. 1956. Differential characteristics *in vitro* and *in vivo* of several substrains of BCG. I. Multiplication and survival *in vitro*. *Am. Rev. Tuberc.* 74:655–66
21. Pierce, C. H., Dubos, R. J. 1956. Differential characteristics *in vitro* and *in vivo* of several substrains of BCG. II. Morphologic characteristics *in vitro* and *in vivo*. *Am. Rev. Tuberc.* 74:667–82
22. Pierce, C. H., Dubos, R. J., Schaefer, W. B. 1956. Differential characteristics

in vitro and *in vivo* of several substrains of BCG. III. Multiplication and survival *in vivo. Am. Rev. Tuberc.* 74: 683–98

23. Dubos, R. J., Pierce, C. H. 1956. Differential characteristics *in vitro* and *in vivo* of several substrains of BCG. IV. Immunizing effectiveness. *Am. Rev. Tuberc.* 74:699–717
24. Guld, J. 1971. Vaccination against tuberculosis. In *Int. Conf. Appl. Vaccines Against Viral, Rickettsial Bact. Dis. Man,* pp. 346–49. Washington DC: Pan Am. Health Organ.
25. Federal Register 1974. 21 CFR, Pt. 630. 39(53):10158-60
26. *WHO Expert Committee on Biological Standardization, 18th Report.* 1966 WHO Tech. Rep. Ser.:No. 329, Annex 1, pp. 25–51
27. Wilson, G. S. 1967. *The Hazards of Immunization,* pp. 66–74. London: Athlone
28. McKhann, C. F., Hendrickson, C. G., Spitler, L. E., Gunnarsson, A., Banerjee, D., Nelson, W. R. 1975. Immunotherapy of melanoma with BCG: Two fatalities following intralesional injection. *Cancer* 35:514–20
29. Mande, R. 1968. Cases of fatal generalized BCG disease. In *BCG Vaccination,* ed. R. Mande, pp. 140–58. London: Dawsons
30. Bergdahl, S., Felländer, M., Robertson, B. 1976. BCG osteomyelitis: Experience in the Stockholm region over the years 1961–1974. *J. Bone Joint Surg.* 58B: 212–16
31. Sutherland, I. 1971. State of the art in immunoprophylaxis in tuberculosis. See Ref. 7, pp. 113–25
32. Stein, S. C., Aronson, J. D. 1953. The occurrence of pulmonary lesions in BCG-vaccinated and unvaccinated persons. *Am. Rev. Tuberc.* 68:695–712
33. Rosenthal, S. R., Loewinsohn, E., Graham, M. L., Liveright, D., Thorne, M. C., Johnson, V. 1961. BCG vaccination against tuberculosis in Chicago: A twenty-year study statistically analyzed. *Pediatrics* 28:622–41
34. 1972. BCG and vole bacillus vaccines in the prevention of tuberculosis in adolescence and early adult life. Fourth report to Medical Research Council by its Tuberculosis Vaccine Clinical Trials Committee. *Bull. WHO* 46:371–85
35. Palmer, C. E., Shaw, L. W., Comstock, G. W. 1958. Community trials of BCG vaccination. *Am. Rev. Tuberc.* 77:877–907
36. Comstock, G. W., Edwards, P. Q. 1972. An American view of BCG vaccination, illustrated by results of a controlled trial in Puerto Rico. *Scand. J. Respir. Dis.* 53:207–17
37. Frimodt-Møller, J., Thomas, J., Parthasanathy, R. 1964. Observations on the protective effect of BCG vaccination in a South Indian rural population. *Bull. WHO* 30:545–74
38. Comstock, G. W., Webster, R. G. 1969. A twenty-year evaluation of BCG vaccination in a school population. *Am. Rev. Respir. Dis.* 100:839–45
39. Comstock, G. W., Palmer, C. E. 1966. Long-term results of BCG vaccination in the southern United States. *Am. Rev. Respir. Dis.* 93:171–83
40. Naganna, K. 1974. Some of the BCG trials and certain aspects involved in them. *Am. Rev. Respir. Dis.* 109:497–99
41. Ferguson, R. G., Simes, A. B. 1949. BCG vaccination of Indian infants in Saskatchewan. *Tubercle* 30:5–11
42. Levine, M. I., Sackett, M. F. 1946. Results of BCG immunization in New York City. *Am. Rev. Tuberc.* 53:517–32
43. Bettag, O. L., Kaluzny, A. A., Morse, D., Radner, D. B. 1964. BCG study at a state school for mentally retarded. *Dis. Chest* 45:503–7
44. Hyge, T. V. 1957. The efficacy of BCG vaccination: Tuberculosis epidemic in a state school with an observation period of 12 years. *Dan. Med. Bull.* 4:13–15
45. Palmer, C. E., Long, M. W. 1966. Effect of infection with atypical mycobacteria on BCG vaccination and tuberculosis. *Am. Rev. Respir. Dis.* 94:553–68
46. Hart, P. D'Arcy. 1967. Efficacy and applicability of mass BCG vaccination in tuberculosis control. *Br. Med. J.* 1:587–92
47. Willis, H. S., Vandiviere, H. M., Vandiviere, M. R., Melvin, I. 1960. Studies in tuberculoimmunity. *Am. J. Med. Sci.* 240:137–57
48. Willis, S., Vandiviere, M. The heterogeneity of BCG. *Am. Rev. Respir. Dis.* 84:288–90
49. Springett, V. H., Sutherland, I. 1970. Comparison of the efficacy of liquid and freeze-dried strains of BCG vaccine in preventing tuberculosis. *Br. Med. J.* 2:148–50
50. Bunch-Christiensen, K., Ladefoged, A., Guld, J. 1968. The virulence of some strains of BCG for golden hamsters. *Bull. WHO* 39:821–28
51. Bunch-Christiensen, K., Ladefoged, A., Guld, J. 1970. The virulence of some

strains of BCG for golden hamsters: further studies. *Bull. WHO* 43:65–70

52. Sirks, J. L., Smith, L., Sekhuis, V. M., Cohen, H. 1974. Comparison of two types of freeze-dried BCG vaccine products. I. Experiments in the laboratory. *J. Biol. Stand.* 2:159–68
53. Sekhuis, V. M., Bleiker, M. A., Sirks, J. L., Cohen, H. 1974. Comparison of two types of freeze-dried BCG vaccine products. II. Results of vaccination of men. *J. Biol. Stand.* 2:169–81
54. Jespersen, A. 1971. *The Potency of BCG Vaccines Determined on Animals,* ed. A. Jespersen. Copenhagen: Statens Seruminstitut
55. Wiegeshaus, E. H., Harding, G., McMurray, D., Grover, A. A., Smith, D. W. 1971. A co-operative evaluation of test systems used to assay tuberculosis vaccines. *Bull. WHO* 45:543–50
56. Guld, J. 1971. BCG as an immunizing agent. See Ref. 7, pp. 149–56
57. Springett, V. H. 1965. The value of BCG vaccination. *Tubercle* 46:76–84
58. 1966. Morbidity and mortality weekly report. Public Health Service recommendations on the use of BCG vaccine in the United States. Atlanta: Cent. Dis. Control 15:350
59. 1975. Morbidity and mortality weekly report. Recommendations of the Public Health Service Advisory Committee on Immunization Practices: BCG vaccines. Atlanta: Cent. Dis. Control 24: 69–70
60. 1974. Infection control in the hospital. *Committee on Infections Within Hospitals.* Chicago: Am. Hosp. Assoc. 3rd ed.

Ann. Rev. Med. 1977. 28:425–52
Copyright © 1977 by Annual Reviews Inc. All rights reserved

THE ASSOCIATION BETWEEN GENES IN THE MAJOR HISTOCOMPATIBILITY COMPLEX AND DISEASE SUSCEPTIBILITY

❖7253

Takehiko Sasazuki, M.D. and Hugh O. McDevitt, M.D.
Division of Immunology, Department of Medicine, Stanford University School of Medicine, Stanford, California 94305

F. Carl Grumet, M.D.
Department of Pathology, Stanford University School of Medicine, Stanford, California 94305

INTRODUCTION

During the past four years, there has been an extraordinarily rapid accumulation of evidence showing that genes of the human major histocompatibility system have a marked effect on susceptibility to a wide variety of diseases. The human major histocompatibility system is a cluster of genes on the VIth human chromosome that determine the structure of cell-surface glycoproteins found on all the cells of the body. These glycoproteins differ from individual to individual and form a complex set of antigenic determinants that constitute the strongest (major) antigenic barrier to tissue transplantation between genetically nonidentical individuals. In addition, by analogy to the major histocompatibility system in several animal species, other genes in the same chromosomal region have marked regulatory effects on immune responsiveness to a wide variety of antigens. Indeed, it appears that almost all of the genes that have been identified in this chromosomal region have some effect upon, or relation to, the immune system, as is shown below.

Perhaps the first demonstration of an association between the major histocompatibility system and susceptibility or resistance to disease was reported by Lilly (1), who showed that genes linked to the mouse major histocompatibility system, the H-2 system, determined resistance to Gross virus–induced leukemia. The mechanism of action of this genetic factor for resistance to viral leukemogenesis was not apparent, but Lilly speculated that it might be due to some type of effect on the immune response to the virus.

In 1967 Amiel (2) reported a moderate increase in the incidence of one particular human transplantation antigenic specificity (at that time designated 4c) in patients with Hodgkin's disease over that in a normal control population. The search for associations between genes in the human major histocompatibility system and disease susceptibility or resistance was given fresh impetus by the demonstration that immune response (*Ir*) genes, controlling the immune response to a wide variety of antigens, were linked to the murine major histocompatibility complex (3). Immune response genes were first described by Benacerraf and his colleagues in studying the immune response of guinea pigs to hapten-substituted synthetic polypeptides such as poly-L-lysine (4). Genetic control of the immune response to other synthetic polypeptides in mice was analyzed by McDevitt (5), in collaboration with Sela and Humphrey.

Ir genes have been reviewed extensively (4–6). The major class of *Ir* genes are those closely linked to the species' major histocompatibility system. These genes control the cellular and humoral response to specific antigens and have been described in the mouse, guinea pig, rat, and monkey. The precise mechanism of *Ir* gene action is not yet known, but they appear to control the ability of immunocompetent lymphocytes to recognize an antigen as foreign, and to initiate both a cellular and humoral response to the antigen. Much of the evidence indicates that *Ir* genes are expressed in thymus-derived lymphocytes (T cells) and may function as a new class of foreign-antigen receptor on the surface of these T lymphocytes. It seems clear that this type of *Ir* gene-controlled antigen recognition does not involve any known immunoglobulins, since none of the structural genes for immunoglobulins have been found to be linked to the major histocompatibility system in any mammalian species. Current evidence [presented in detail in (6)] indicates that genes mapping in the *I*, or immune response region of the mouse major histocompatibility system (see below), determine the structure and/or function of cell-surface receptors for foreign antigens on the surface of T lymphocytes, and possibly also on the surface of macrophages and B lymphocytes (the precursors of antibody-producing cells).

The demonstration that genes in the major histocompatibility system have a profound effect on determining the ability to respond well or poorly to specific antigens suggested that genes in the major histocompatibility system might have a major effect on resistance to those diseases in which the immune system was protective, or in autoimmune diseases, in which the immune system itself might initiate the disease process. The possibility of such a functional relationship, as well as the earlier demonstrations of associations between the major histocompatibility system and disease, led to a series of studies in many laboratories around the world, searching for associations between the human major histocompatibility system, the *h*uman *l*eukocyte *a*ntigen (HLA) system—and disease susceptibility.

During the past four years, this search has been extraordinarily productive and has led to the discovery that a number of rheumatic, autoimmune, neurologic, endocrine, gastrointestinal, and dermatologic diseases show definite associations with particular genetic variants (alleles) of several genes in the human major histocompatibility system. Before discussing the human major histocompatibility system and the disease associations in detail, a brief description of the genetic structure of the murine major histocompatibility system—the H-2 system—will permit us to indicate some of the functional relationships between this complex genetic region and the immune system.

THE H-2 SYSTEM

Figure 1 is a schematic diagram of the H-2 system. [For a detailed discussion of the genetic fine structure of the murine H-2 system, its gene products, and some of their functions, the reader is referred to a comprehensive review by Shreffler & David (7).] The H-2 system is bounded by *H-2K* and *H-2D*, which are the initially described, classical transplantation antigens that can be detected by cytotoxic T cells as well as by alloantisera. The *H-2K* and *H-2D* gene products are 45,000-MW cell-surface glycoproteins found on all the tissues of the body, with the possible exception of sperm and very early embryonic tissues. The human counterparts are the *HLA-B* and *-A* genes, respectively (see Figure 2).

The true function of the *H-2K* and *H-2D* gene products is not known; however, recently it has become apparent that these genes are intimately involved in the induction and effector phases of T-cell–mediated cytotoxicity, or T-cell killer function. In several systems it has been shown that cytotoxic T cells specific for viruses, and for minor transplantation antigens such as the H-Y or male-transplantation antigen, recognize not only the antigen to which they have been sensitized, but also the *H-2K* and/or *D* gene product on the immunizing cells. Thus, after an animal is immunized with foreign cells carrying a particular virus, its T cells will only recognize virus on the surface of cells of the same foreign transplantation antigen type as those originally used for immunization. Even though the animal has been

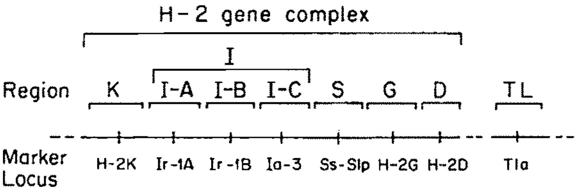


Figure 1 A schematic diagram of the H-2 complex. The H-2 system is bounded by the *H-2K* and *H-2D* loci. The *I* region of the H-2 system includes *Ir-1A*, *Ir-1B*, and *Ir-1C*. The corresponding Ia regions are *I-A*, *I-B*, and *I-C*. Earlier genetic maps of the *I* region included only the *Ir-1A*, *B*, and *C* regions, but recent analysis of several *H-2* recombinants has indicated that a new region, *I-J*, must be intercalated between *Ir-1B* and *Ir-1C*, and that another region exists between *I-J* and *Ir-1C*, which has been given the preliminary designation of *I-E* (13). The *S* locus or region of the H-2 system determines at least two genes, one of which determines the structure of a component of the C4 molecule. There is also evidence that the *S* region determines the levels of C1, 2, and 3.

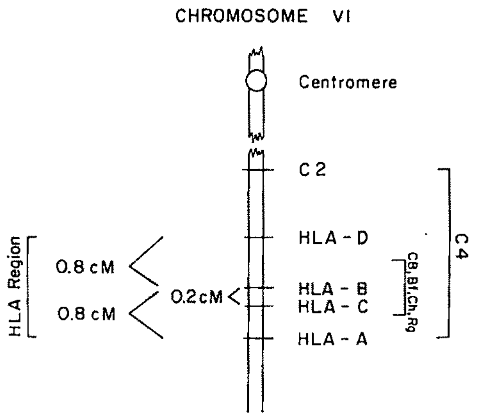


Figure 2 A schematic diagram of the HLA system on human chromosome VI. *HLA-B* and *A*, which are analogous to *H-2K* and *D*, are separated by 0.8 recombination units or centimorgans (cM). The *HLA-C* locus does not have a counterpart in the mouse. The *HLA-D* locus or region determines the structure of lymphocyte cell-surface antigens eliciting the mixed-culture reaction, and therefore may be analogous to the *I* region of the mouse. The genes for C4 deficiency, C8 deficiency, Bf (properdin factor B) and for the Chido (Ch) and Rogers blood-group antigen have not been precisely mapped, but are within the *HLA* system.

immunized to these virus-bearing cells, it will not kill cells of its own transplantation antigen type that carry the same amount of virus on their cell surface (8). It is thus clear that the *H-2K* and *H-2D* gene products have a major effect on the development and specificity of cell-mediated killer functions. It is not yet known whether this is the primary or even a major function of these gene products. There has been considerable speculation that these gene products, because of their ubiquitous tissue distribution, must play a more fundamental role in cellular interaction and tissue morphogenesis. As yet, however, there is no definite evidence for these latter postulated functions.

The *I* region was originally defined because genetic control of specific immune response to several polypeptides was shown to map in the chromosomal region lying between the *H-2K* and *S* genes of the H-2 system (9). Within a short time, it became apparent that in addition to genetic control of specific immune responses, this region also determined the structure of cell-surface antigens that elicited the mixed-lymphocyte-culture reaction and the graft-versus-host reaction. Studies designed to produce antisera against *I*-region gene products resulted in the identification of a new class of cell-surface alloantigens, the *I*mmune-response region–*a*ssociated antigens, or Ia antigens. These cell-surface allo-antigens are also glycoproteins that fall into two molecular-weight ranges, 25,000 and 33,000, and are found predominantly on the surface of lymphocytes, macrophages, sperm, and epidermal cells. A systematic analysis of a number of different inbred mouse strains, and of recombinant H-2 chromosomes derived by crossing over between the *H-2K* and *H-2D* loci, led to the subdivision of the *I* region into the *I-A*, *I-B*, and *I-C* subregions. Genes controlling the immune response to particular foreign antigens were mapped in the *I-A*, *I-B*, or *I-C* regions and were designated *Ir-1A, Ir-1B,* and *Ir-1C.* In several instances it also became apparent that two *I*-region genes were required to develop a high response to a particular antigen (10).

It is now apparent that the Ia antigens are the primary cell-surface alloantigens that stimulate in the mixed-lymphocyte-culture reaction in vitro and that elicit the graft-versus-host reaction in vivo. A major research effort now centers on determining the relationship of the Ia antigens to *Ir* gene function. During the past year, it has been shown that factors can be isolated from thymus-derived lymphocytes that appear to be capable of regulating specific immune responses. These factors mimic the effects of helper thymus-derived lymphocytes and thus potentiate the immune response to a particular antigen (11), or they mimic the effects of suppressor thymus-derived lymphocytes (12) and suppress the specific immune response to a particular antigen.

Both types of factor have a number of interesting properties in common. They have a molecular weight in the 40,000–50,000 range; they bind specifically to antigen-immunoabsorbent columns; they lack all immunoglobulin antigenic determinants; and they possess Ia antigenic specificities. Preliminary evidence indicates that helper factors bear Ia determinants mapping in the *I-A* subregion (12) and suppressor-factor Ia determinants map in a new *I* subregion that has tentatively been placed between *I-B* and *I-C*, and has been given the preliminary designation of *I-J* (13). These preliminary findings are important because they indicate that Ia antigens may be part of structures derived from T cells that have the ability to recognize specific antigenic determinants. If these results are borne out, they indicate that the *I* subregion of the H-2 system contains a number of genes that are in some way involved in the structure of a new class of antigen-specific receptors found on the surface of immunocompetent T lymphocytes.

Ir genes have not yet been conclusively demonstrated in man, so that a precise human counterpart to the murine *I* region does not yet exist. However, the Ia antigens elicit the mixed-lymphocyte reaction [MLR (see below)] and lymphocyte antigens determined by the *HLA-D* locus have a similar function (see Figure 2). By analogy, it is presumed that the *HLA-D* region may be the human *I* region.

The major histocompatibility system has yet another intimate relationship with the immune system. As is shown below, structural genes for several complement components (the second, fourth, and eighth complement components) are linked to the human major histocompatibility system, the HLA system. In the mouse, recent studies (7) have shown that the *S* gene of the H-2 system, which had been known to control the structure and serum levels of a serum β-globulin, is in fact the structural gene for one of the polypeptide chains of the fourth component of complement C4 (7). (The human C4 gene has not been accurately mapped as yet, cf Figure 2.) Thus, in addition to being intimately involved in the specificity of T-cell–mediated cytotoxicity, and in regulating both the potentiation (help) and the inhibition (suppression) of the immune response, the major histocompatibility system also determines the structure of several complement components.

It is not yet known whether the close linkage of all of these genes, which regulate cellular, humoral, and effector arms of the immune response, indicates a functional interrelationship and an underlying structural homology between these three genetic regions of the major histocompatibility system. However, the extraordinary finding that all three classes (the *K* and *D* genes, the *I*-region genes, and the *S*-region genes) of the major histocompatibility system are involved in particular functional aspects of the immune response suggests that regulation of the immune response is a major function of the major histocompatibility system and that the different genes in this complex system may be both functionally and structurally interrelated (14).

In many ways, the H-2 system has served as a prototype for extending our understanding of the human major histocompatibility system, the HLA system. This system is described in detail in the following section.

THE HLA SYSTEM

Background

Studies of the human major histocompatibility (HLA) system began in the late 1950s, when it was reported that leukoagglutinating antibodies could be detected in the sera of multiply transfused patients, and also of parous women (15, 16). Because these antibodies reacted with some, but not all, leukocytes from normal individuals, it was recognized that the leukocyte antigens involved belonged to a genetic polymorphic system. As many more such sera became identified during the early 1960s, studies of families, and computer-assisted analyses of population studies with these antisera, provided the first examples of the allelic nature of the antigens detected (17–19), i.e. that these antigens were variant, or allelic, products of the same gene (or genes). During that same time, it was shown that the leukocyte antigens could also be detected by lymphocyte cytotoxicity.

The development of a microtechnique soon led to the general acceptance of the antibody- and complement-mediated lymphocyte microcytotoxicity assay as the standard in the field (20). This technique, which has changed little over the succeeding decade, basically consists of a 1–2-hr incubation of approximately 2000 viable lymphocytes with each of a selected series of antisera, plus fresh-frozen rabbit serum as a source of complement. At the end of the incubation period, each test is read for residual viable cells using a vital dye. Although additional techniques have been developed, the antibody-mediated, lymphocyte microcytotoxicity technique remains the most widespread and accepted standard. As data concerning this system began to accumulate, the need for closer collaboration among investigators in the field was realized and a series of international workshops, held approximately every two years, was initiated in 1964. A well-documented model of the very complex genetic system we now refer to as HLA emerged from those workshops, as well as from a great deal of work carried on in the laboratories of many independent investigators. [For general reviews see (21–23).]

Serologically Detectable Antigens

The genetic determinants of all the HLA antigens reside in a segment of chromosome VI that is estimated to be approximately 1000 genes long and is referred to as the HLA region (21–27). Within this region lie the four currently defined HLA loci, *HLA-A, HLA-B, HLA-C,* and *HLA-D.* Antigens determined by the *A, B,* and *C* loci are present on the surfaces of all nucleated cells (also on blood platelets) and are detectable by serological techniques such as the lymphocyte cytotoxicity method described above. Each locus has its own series of unique antigens that are the products of the allelic genes belonging to that particular locus. Antigens that can be identified by presently available antisera are shown in Table 1. Antigens written with an A, B, or C followed by an arabic numeral are those acknowledged by the International Workshops and the World Health Organization to be clearly defined. Those written with a w preceding the arabic numeral are less clearly defined and are the subject of further workshop evaluation, whereas those antigens without an A, B, or C prefix have been reported by only a few laboratories and remain to be confirmed at an International Workshop. There are currently 19 well-characterized *A* locus antigens and 24 *B* locus antigens, accounting for over 95% of the *A* and *B* locus genes (28, 29). The more recently defined *C* locus has only six antigens agreed upon, accounting for only about half of the alleles at this locus.

HLA typings are written as in Table 2. Each individual possesses two HLA regions, one paternally and one maternally derived. Because the HLA antigens are

Table 1 HLA antigens

A Locus	B Locus	C Locus	D Locus
HLA-A1	HLA-B5		
-A2	-B7		
-A3	-B8		
-A9	-B12		
-A10	-B13		
-A11	-B14		
-A28	-B18		
-A29	-B27		
HLA-Aw23	HLA-Bw15	HLA-Cw1	HLA-Dw1
-Aw24	-Bw16	-Cw2	-Dw2
-Aw25	-Bw17	-Cw3	-Dw3
-Aw26	-Bw21	-Cw4	-Dw4
-Aw30	-Bw22	-Cw5	-Dw5
-Aw31	-Bw35		-Dw6
-Aw32	-Bw37		
-Aw33	-Bw38	-T7	
	-Bw39		
-Aw36	-Bw40		-LD-107
-Aw43	-Bw41		-LD-108
	-Bw42		
	TT		
	KSO		
	Hs		
	407*		

codominantly expressed and there are no known silent or amorphic alleles, each individual is characterized by two antigens from each of the loci (i.e. a set of six serologically determined antigens). Because of the close linkage between the loci, the entire HLA region is inherited intact in approximately 99% of offspring. Within any given family, then, a particular haplotype (i.e. a fixed combination of HLA antigens determined by the HLA region of a single chromosome VI) will be inherited so that a specific set of A, B, and C antigens will always appear together. Because the crossover frequency is finite, however, hundreds of generations of random human matings have resulted in every possible combination of *A, B,* and *C* genes. In the random population, therefore, the presence of a particular antigen at one of the loci generally does not affect the probability of finding any particular antigens at the other loci.

An illustration of the transmission of HLA haplotypes is given in the pedigree shown in Table 3. In six of the seven children, the haplotypes have been inherited intact as would be expected. One child (#6), however, demonstrates a rare (<1% of offspring) event defined as an HLA crossover. During meiosis a recombination between maternal chromosomes has occured so that a new haplotype has been

Table 2 Common genetic terminology

Haplotype	=	A combination of antigens determined by the HLA region of a single chromosome, e.g. A1, B8, Cw3, Dw5
Genotype	=	A pair of haplotypes possessed by a particular individual, e.g. A1, B8, Cw3, Dw5; A2, B12, Cw4, Dw3
Phenotype	=	The antigens possessed by a particular individual regardless of haplotype assignment e.g. A1, A2; B8, B12; Cw3, Cw4; Dw3, Dw5

Table 3 Transmission of HLA haplotypes in a family

Individual	HLA genotype
Father (a/b)	A1, B8, C–, D–; A3, B18, C–, D–
Mother (c/d)	A2, B27, Cw1, Dw1; A29, B12, C–, D–
Child #1 (a/d)	A1, B8, C–, D–; A29, B12, C–, D–
Children #2, 3, 5, 7 (b/d)	A3, B18, C–, D–; A29, B12, C–, D–
Child #4 (b/c)	A3, B18, C–, D–; A2, B27, Cw1, Dw1
Child #6 (b/cd)[a]	A3, B18, C–, D–; A2, B12, C–, D–
	c d

[a] Child #6 possesses the recombinant chromosome (cd). The designations C– and D– represent as yet unidentified antigens at the *C* and *D* loci. Data kindly provided by Dr. Rose Payne, Department of Medicine/Division of Hematology, Stanford University School of Medicine, Stanford, California 94305).

created. The A2 from the c chromosome is now combined with the B12, C-, and D- of the d chromosome. This new haplotype can now be passed on intact to future descendants of child #6.

As with other polymorphic genetic markers, the frequency of each antigen will depend upon the ethnic group studied (28). For example, the HLA-A1 antigen appears in 33% of Caucasians, but is absent from Chinese, whereas the Hs antigen occurs in 25% of Chinese, but is absent from Caucasians (30). This marked ethnic variability is, of course, an important consideration in selecting appropriate controls for any population study utilizing the HLA system.

The chemical structure of the HLA antigens is an area of active research at present. The serologically detectable antigens have been solubilized from the cell surface and consist of a dimer with a heavy glycoprotein chain (MW 44,000) containing the antigenic specificity, and a β_2 microglobulin chain (MW 12,000) (31). The β_2 microglobulin does not possess HLA antigenic markers and the locus for the β_2 chain has been located on chromosome XV (32).

The D Locus and Mixed Leukocyte Cultures (MLC)

Unlike the serologically defined A, B, and C antigens, *D*-locus markers can at present be recognized only by an in vitro lymphocyte culture test. Although some variations of technique are now used, the basic test consists of several days of culture of lymphocytes of one individual (the responder) with X-irradiated or mitomycin-C–treated lymphocytes of a second individual (the stimulator). If the stimulator possesses *D*-locus antigens not present in the responder, the lymphocytes of the latter undergo transformation, resulting in DNA synthesis that can be quantified as an index of response. (The X-irradiation or mitomycin-C prevents the stimulator cells from reacting similarly.) Through careful selection of individuals homozygous for *D*-locus alleles (e.g. offspring of first-cousin marriages), it is possible to obtain "typing cells," which may then be used to test others for the presence or absence of a particular *D*-locus antigen (33). *D*-locus typing is technically difficult and has only recently been developed, so that only six allelic types are generally agreed upon for this locus (34). Because MLC nonstimulation between random individuals is such a rare event, it is anticipated that many *D*-locus alleles exist and are yet to be identified. The problem is further complicated by the likelihood that the *D* locus is several (or many) loci, as is the case for the analogous *I* region in the mouse.

Linked Loci and Linkage Disequilibrium

Through the studies of families with recombinant haplotypes, it has been possible to map not only the *HLA* loci, but also a number of other loci as well. A graphic presentation of the HLA region is shown in Figure 2. It is particularly interesting to note that three components of the complement system are linked to the *HLA* loci, thus defining yet another important function (in addition to transplantation antigenicity, disease susceptibility linkage, and probably human immune-response genes) for this segment of chromosome VI.

Based on the principle of random human matings over hundreds of generations, with small but finite recombination frequencies within the region, almost all possible haplotype combinations should appear at a frequency proportional to the product of the frequencies of the individual component genes. Although this rule is generally valid, some exceptions do occur such that certain combinations of alleles at different loci appear together on the same chromosome more frequently than would be predicted. When this occurs, the genes are said to be in linkage disequilibrium. For example, in Danes, the frequency of the haplotype containing both the A1 and B8 antigens is 0.099, whereas its predicted frequency (if no linkage disequilibrium existed between those two antigens) is 0.022 (23). Among Caucasians of the Middle East or Asia, however, the linkage disequilibrium between those same two antigens is not significant, i.e. A1 and B8 do not occur together more frequently than predicted by random matings.

No mechanism has yet been proven to account for the observed linkage disequilibrium between HLA antigens; however, several may be postulated. For example, some unknown survival value might accrue in some environments from certain combinations of antigens, some combinations might affect chromosomal stability and therefore have relatively lower recombination frequencies, or some combinations might derive from relatively new mutants that have not yet had time to dissociate to a state of linkage equilibrium.

Regardless of mechanism, the phenomenon of linkage disequilibrium exists sufficiently frequently in the HLA system to suggest that some biological function may underlie the associations observed (see below).

B-Cell Antigens

In several animal species the existence of immune response (*Ir*) genes linked to the species' major histocompatibility region has been clearly documented, and preliminary evidence to support the existence of similar genes in man has also been gathered (5, 35). In the best-studied model, that of the mouse, loci for the *Ir* genes map within the *H-2* region, closer to the *K* than to the *D* locus. Also mapping at the same (or as yet genetically indistinguishable) loci as the *Ir* genes are a series of immune response–associated *Ia* genes (36). The products of the *Ia* genes are antigens present in greatest concentration on the surface of B lymphocytes (and also in small quantities on T lymphocytes) and detectable by antibody-mediated lymphocyte cytotoxicity.

The exact function of Ia antigens and their relation to the functionally defined *Ir* genes is not yet known. It is reasonable to postulate that Ia markers may represent the serologically detectable *Ir* gene cell-surface receptors for antigen recognition. For this reason, a great deal of research activity has recently centered upon a search for human B-cell antigens as possible murine Ia analogs [several such antigens have already been described (37–39)] and as possible serologically detectable *D* locus gene products.

A number of methods for testing for human B-cell antibodies are available, and one of the most common is to first remove HLA-A, B, or C antibodies from antisera by their absorption onto platelets (which lack "Ia" antigens). The absorbed sera are then tested for residual antibodies against purified B cells obtained either from patients with chronic lymphocytic leukemia (a disease characterized by proliferation of a clone of B cells) or from normal B cells separated from T cells by virtue of differential adherence to sheep cells or anti-immunoglobulin columns. Preliminary studies reporting the identification of B-cell antigens in man have already been published, and the next (1977) International HLA Workshop will focus on this aspect of HLA serology, particularly the clarification of B-cell antigens and their genetic mapping and probable correlations with *D*-locus genes.

SPECIFIC ASSOCIATIONS BETWEEN HLA AND DISEASE

The first association between HLA and disease was reported by Amiel in 1967 (2) involving Hodgkin's disease and an *HLA-B* locus antigen then designated 4c. Over the next 10 years a progressively larger number of diseases has revealed association or linkage with HLA-A, B, or C antigens. The most striking of these is still the association between B27 and ankylosing spondylitis (AS), an association so strong that it has been utilized as an aid to clinical diagnosis.

In 1973, Jersild et al (40) reported the first association of a disease with a *D*-locus antigen, multiple sclerosis and Dw2. These findings suggested that the less impressive, but earlier reported association of HLA-B7 with multiple sclerosis was possibly secondary to the existence of strong linkage disequilibrium between B7 and Dw2. More recently Terasaki et al (41) reported an even stronger association of multiple sclerosis with a serologically defined B-cell alloantigen type, Group II, which is linked to HLA and may be related to the HLA-D antigen Dw2. As finer genetic mapping of the HLA region evolves, it may be possible to define even closer markers of disease-susceptibility genes, and eventually even to test directly for the disease-susceptibility genes themselves.

Although rapid progress is being made in analyzing other HLA and disease associations, as is shown below, a number of precautions must be observed in studying these associations. First, proper control populations must be selected for each new patient population, particularly with respect to racial background. Second, diagnostic classification is critical. For example, if diabetes mellitus was taken as a single disease, it would not have been possible to detect an association with any HLA type. Only when juvenile-onset, insulin-dependent diabetes was considered separately was an association with HLA-B8 evident. Similarly, if psoriasis vulgaris is not differentiated from pustulosis palmaris et plantaris, the associations of B13, Bw17, and Bw37 with the former disease would be masked or diluted out. This type of problem is called "dilution" and may obscure a true association.

Third, quantitative measures of the strength of HLA and disease associations must be considered carefully. An accepted measure of the intensity of an association is the relative risk (RR):

$$\mathrm{RR} = \frac{(\#\text{ patients with the HLA antigen}) \times (\#\text{ controls without the antigen})}{(\#\text{ controls with the antigen}) \times (\#\text{ patients without the antigen})}$$

a calculated term that represents the risk of developing a disease in an individual possessing a particular HLA antigen as compared to an individual lacking that antigen. In determining the statistical significance of the associations it is also important to be aware that p values (or X^2) may be misleading unless the number of variables examined in any comparison of populations is also taken into account. Thus, if one is studying the association of 20 different *HLA-B* alleles with a given disease, it is likely that one of these 20 will show a significant difference (uncorrected $p < 0.05$) in frequency from the control population by chance alone. Multiplying an uncorrected p value by the number of HLA antigens tested for provides a roughly "corrected p," and ensures that probability values observed are not artificially elevated simply because of the large number of variables studied.

Finally, a distinction must be made between association and linkage. Association is demonstrated in population studies of unrelated individuals when two particular traits (e.g. a selected HLA antigen and expression of a specific disease) occur together at a frequency different from that predicted by chance alone. Linkage cannot be proven by population studies, but rather must be shown by studies in families in which segregation of two traits (e.g. an HLA antigen and C4 deficiency, or an HLA antigen and a specific disease) can be demonstrated in a consistent pattern. Linked traits do not necessarily show association, and vice versa.

Despite the difficulties of performing and analyzing population studies, approximately 40 diseases have already been shown to be associated with HLA antigens, and several others demonstrate linkage. The majority of these associations, arranged into categories according to medical specialty, are listed in Tables 4–10 and are briefly discussed below. [Because of space limitation, primary references have been given only for those diseases not already cited in (42) and (43).]

Rheumatology (Table 4)

The strongest HLA and disease association observed is that found between AS and HLA-B27. This association is so strong that B27 typing has become a clinically useful aid in the diagnosis of this disease at an early stage, i.e. prior to the appearance of characteristic X-ray findings. More than 90% of Caucasian patients with AS possess B27, whereas only 5–9% of controls have this specificity. This strong association is also prominent in other ethnic groups. For example, among some Canadian Indians, 100% of AS patients had B27. These data are significant even though the number of patients tested was small and the normal frequency of B27 in this ethnic group is extremely high compared with that of Caucasians. In Japanese, in whom B27 is extremely rare, AS again has a strong association with B27.

Population surveys have revealed that roughly 20% of the B27(+) population has radiologically detectable sacroiliitis, which is frequently symptomatic and previously undetected (44). The sex incidence is equal in these undetected cases. Thus, some manifestations of AS are ten times more frequent than previously thought in males, and many, many times more frequent than previously thought in females. The total number of such cases in the US population would thus approach two million. [250×10^6 population $\times$ 0.05 B27 (+) frequency $\times$ 0.20 frequency of AS findings in B27(+) individuals.]

Table 4 Rheumatology

Disease	HLA antigen	% Antigen frequency patient	% Antigen frequency control	Relative risk	p Value
Ankylosing spondylitis					
Caucasians	B27	90	8	87.8	1.0×10^{-10}
Haida Indians	B27	100	51	34.4	0.17×10^{-4}
Bella Coola Indians	B27	100	26	20.2	0.19×10^{-1}
Japanese	B27	67	0	305.7	1.0×10^{-10}
Reiter's syndrome	B27	78	8	35.9	1.0×10^{-10}
Yersinia arthritis	B27	79	9	24.3	1.0×10^{-10}
Salmonella arthritis	B27	67	9	17.6	0.36×10^{-9}
Psoriasis arthropatica					
peripheral	B13	10	6	2.2	0.95×10^{-2}
	B27	16	9	2.5	0.35×10^{-3}
	Bw17	25	6	5.8	1.0×10^{-10}
	Bw38	13	3	4.5	0.32×10^{-3}
central	B13	20	6	4.8	0.76×10^{-7}
	B27	40	9	8.6	1.0×10^{-4}
	Bw17	12	6	2.5	0.46×10^{-2}
	Bw38	23	3	9.1	0.56×10^{-6}
Psoriasis arthropatica					
unspecified	B27	42	8	7.1	0.25×10^{-9}
	Bw17	25	8	3.9	0.73×10^{-3}
Juvenile					
rheumatoid arthritis	B27	26	9	4.7	1.0×10^{-10}
Rheumatoid arthritis	Dw4	36	16	3.0	0.5×10^{-2}

Reiter's syndrome, defined strictly by the triad of arthritis, nongonococcal urethritis, and conjunctivitis, also has a strong association with B27. Available evidence suggests that this syndrome can appear as a complication of Shigellosis (flexneri) occurring primarily in B27-positive persons (45, 46). It is also of interest to note that patients with gonococcal arthritis rarely possess the B27 antigens, raising questions about an etiologic relationship between gonococci and Reiter's syndrome (47, 48).

Reactive arthritis following exposure to infectious agents such as *Yersinia enterocolitica* and *Salmonella* has also shown an association with B27. Psoriatic arthritis on the other hand, has a weaker association with B27 and significant associations with B13, Bw17, and Bw37. These latter associations are similar to those of all psoriasis vulgaris patients, as seen in Table 6. Juvenile rheumatoid arthritis (JRA) also has a weaker association with B27. The observation that each of these types of arthritis is associated with the same HLA antigen, B27, strongly suggests that they may share some combination of common determinants of susceptibility, pathogens, or pathogenic process.

Unlike the preceding forms of arthritis, classical adult rheumatoid arthritis (RA) is not associated with B27, but rather is associated with the HLA-D specificity Dw4 (or R, r) (49, 50). The absence of an association of RA with B27 raises the possibility that this disease is pathogenetically distinct from the B27-associated forms of arthritis. RA is the first disease to show an association only with an *HLA-D* locus type, in the absence of any *HLA-A, B,* or *C* association. This is potentially a very important observation, and is discussed in the section on Mechanisms.

Rheumatic fever and gout have not shown any significant associations.

Neurology—Psychiatry (Table 5)

Multiple sclerosis (MS) was first reported to have a slight but significant association with HLA-A3, and later, a stronger association with B7. Subsequently, HLA-Dw2 was shown to be even more strongly associated. These data suggest that the *D* locus may be the closest of the *HLA* loci to the actual locus determining MS susceptibility, and that the disease associations of B7 and even A3 may merely arise because of the linkage disequilibrium of B7 with Dw2, and of A3 with B7. More recently, a B-lymphocyte alloantigen has been shown to be more strongly associated with multiple sclerosis than Dw2, suggesting that the locus for this new B-cell alloantigen may be yet closer to the actual MS disease-susceptibility locus.

Myasthenia gravis (MG) has a strong association with B8, and a weaker association with A1, presumably secondary to the strong linkage disequilibrium between A1 and B8. In a preliminary study of a small group, an association of MG with Dw3 also appears (30% of patients versus 9% of controls). Manic-depression has been reported to be associated with Bw16, and for both schizophrenia and paralytic poliomyelitis conflicting reports have appeared describing the presence or absence of HLA associations. Each of these diseases must be further studied before any conclusion should be drawn regarding the validity of their HLA associations.

Table 5 Neurology

Disease	HLA antigen	% Antigen frequency		Relative risk	*p* Value
		patient	control		
Multiple sclerosis	A3	36	26	1.5	0.94×10^{-8}
	B7	34	25	1.6	0.76×10^{-9}
	Dw2	67	18	6.9	0.17×10^{-5}
	Group 4[a]	84	33		0.3×10^{-2}
Myasthenia gravis	A1	45	26	2.5	0.87×10^{-9}
	B8	52	23	4.4	1.0×10^{-10}
Paralytic poliomyelitis	B7	21	21	1.0	0.65×10^{-3}
	B7	38	19	2.6	
Schizophrenia	A28	19	6	3.5	0.28×10^{-5}
	A28	4	7	0.7	
Manic-depression	Bw16	13	5	2.8	0.79×10^{-4}

[a] Proposed B-cell alloantigen

Dermatology (Table 6)

Psoriasis vulgaris has an association with three different *HLA-B* locus antigens, B13, Bw17, and Bw37. Psoriasis (nonspecific) is associated with Bw16, as well as with B13 and Bw17. Psoriasis pustulosis, on the other hand, has been associated with B27.

Dermatitis herpetiformis (DH) is associated with Dw3, B8, and A1. As with some other HLA and disease associations involving more than one HLA antigen, the primary association of DH may be with the B8 antigen, and the weaker association with A1 may arise as a result of the linkage disequilibrium between B8 and A1. More recently, it has been reported that an antiserum obtained from the parous wife of a patient with dermatitis herpetiformis reacted with the B lymphocytes from 15 of 15 patients with this disorder, but none of 37 healthy controls (51). The relation of this apparent B-cell–specific antigen to the HLA region is not yet known.

Behçet's disease, rare in both Europe and America, but fairly common in the eastern Mediterranean and in northern parts of Japan, has an association with B5 in both Caucasians and Japanese. Recrudescent herpes labialis has been reported to be associated with A1 and B8, and pemphigus is associated with A10. Scleroderma, acne conglobata and recurrent aphthous stomatitis show no HLA associations in Caucasians, and other ethnic groups have not yet been studied.

Table 6 Dermatology

Disease	HLA antigen	% Antigen frequency		Relative risk	p Value
		patient	control		
Psoriasis vulgaris	B13	20	5	4.7	1.0×10^{-10}
	Bw17	26	8	5.0	1.0×10^{-10}
	Bw37	8	1	6.4	0.73×10^{-7}
Psoriasis nonspecificata	B13	15	5	3.9	1.0×10^{-10}
	Bw16	16	3	4.3	0.21×10^{-10}
	Bw17	28	7	5.4	1.0×10^{-10}
Psoriasis pustulosis	B27	28	9	3.8	0.18×10^{-3}
Dermatitis herpetiformis	A1	69	30	4.4	1.0×10^{-10}
	B8	77	25	9.2	1.0×10^{-10}
	A10	39	13	3.1	0.12×10^{-3}
Behçet's disease					
Caucasian	B5	35	11	4.3	0.81×10^{-4}
Japanese	B5	75	31	6.5	0.24×10^{-5}
Recrudescent herpes labialis	A1	56	25	3.7	0.13×10^{-7}
	B8	33	17	2.5	0.35×10^{-3}

Endocrinology (Table 7)

Juvenile-onset, insulin-dependent diabetes is associated with B8 and Bw15. It is also interesting that both B8- and Bw15-positive persons have a much higher relative risk than persons with only B8 or Bw15, but individuals homozygous for B8 or Bw15 do not differ in relative risk from individuals heterozygous for B8 and Bw15. These data indicate that there is no gene dose effect for either B8 or Bw15 alone, but that the two genes together yield a complementary or additive effect and therefore each may affect disease susceptibility in a different manner. No significant deviation of HLA specificities was found in maturity-onset diabetes mellitus. Because one subgroup of diabetics (juvenile-onset) are characterized by an increased frequency of a specific genetic marker (B8 and/or Bw15), and another subgroup (adult-onset) is not, it is entirely possible that the syndrome of diabetes may arise from distinct pathogenetic mechanisms in each of these disease subgroups. Further subdivision of diabetes into even more homogeneous groups might yield still additional associations, suggesting the existence of still other pathogenetic mechanisms.

Thyrotoxicosis, particularly Graves's disease, is associated with B8 in Caucasian and with Bw35 in Japanese patients.

Subacute thyroiditis (de Quervain type) also has an association with Bw35 in Caucasians. Hashimoto's disease, on the other hand, lacks any significant HLA associations.

Idiopathic Addison's disease has an association with Dw3 and also with B8, an antigen that is in linkage disequilibrium with Dw3.

It should also be noted that several autoimmune diseases with organ-specific autoantibodies, such as juvenile-onset, insulin-dependent diabetes, Graves's disease, Addison's disease, and MG each have associations with B8 and many have even

Table 7 Endocrinology

Disease	HLA antigen	% Antigen frequency		Relative risk	p Value
		patient	control		
Diabetes (juvenile-onset, insulin-dependent only)	B8	37	22	2.1	1.0×10^{-10}
	Bw15	23	15	2.1	1.0×10^{-10}
	Dw3	50	18	4.5	0.7×10^{-2}
Thyrotoxicosis					
Caucasian	B8	42	24	2.5	1.0×10^{-10}
Japanese	Bw35	57	21	5.0	0.46×10^{-4}
Addison's disease	B8	50	23	3.9	0.73×10^{-5}
	Dw3	70	18	10.5	0.3×10^{-2}
Subacute thyroiditis (de Quervain type)	Bw35	77	13	22.2	1.0×10^{-10}

stronger associations with Dw3. These findings suggest the possibility that these syndromes may share some common step in their pathogenesis.

Gastroenterology (Table 8)

Chronic active hepatitis is associated with B8 and A1, with the latter antigen association possibly due to the linkage disequilibrium between A1 and B8. An increased frequency of homozygosity for B8 was reported (28.5% in the patient group compared to 2.8% in the control group), indicating that the susceptibility gene for chronic active hepatitis might be recessive with linkage disequilibrium with B8. It was also reported that healthy carriers of the hepatitis-associated antigen HB_sAg have an increased frequency of Bw41.

Idiopathic hemochromatosis, a disease with an increased familial incidence, but without any clear-cut immunologic abnormalities in its pathogenesis, has been associated with A3 and B14. These antigens are not in linkage disequilibrium with each other.

Celiac disease has strong associations with A1 and B8. The relative risk for Bw35 in this disease is substantially less than 1.0 (0.47), indicating a possible disease resistance in Bw35 persons.

Ulcerative colitis and Crohn's disease are reported to have no significant associations with any HLA specificities, and pernicious anemia has had conflicting reports relating to a weak association with B7.

Oncology (Table 9)

Hodgkin's disease was the first reported association of an HLA specificity and disease. In the ten years subsequent to that report, intensive studies have been done

Table 8 Gastroenterology

Disease	HLA antigen	% Antigen frequency		Relative risk	p Value
		patient	control		
Chronic active hepatitis	A1	42	28	1.8	0.64×10^{-4}
	B8	44	20	3.0	1.0×10^{-10}
Idiopathic hemochromatosis	A3	78	27	9.5	0.21×10^{-10}
	B14	26	3	9.2	0.53×10^{-5}
Celiac disease	A1	64	30	4.2	1.0×10^{-10}
	B8	71	23	8.9	1.0×10^{-10}
	Bw35	7	17	0.5	0.39×10^{-3}
HB_sAg	Bw41	12	1	11.2	0.72×10^{-3}
Pernicious anemia	B7	36	22	2.2	0.12×10^{-3}

Table 9 Oncology

Disease	HLA antigen	% Antigen frequency patient	 control	Relative risk	p Value
Hodgkin's disease	A1	39	31	1.4	1.0×10^{-6}
	B5	16	11	1.6	1.0×10^{-6}
	B8	29	24	1.3	1.0×10^{-4}
	B18	13	7	1.9	1.0×10^{-6}
Acute lymphocytic leukemia	A2	60	54	1.3	1.0×10^{-2}
Retinoblastoma	B12	10	25	0.3	0.27×10^{-3}
	Bw35	25	11	2.8	0.38×10^{-3}
Nasopharyngeal cancer	Hs(Sin-2)	44	21	3.0	0.7×10^{-2}
	Sin-2a[a]	100	10	–	0.1×10^{-2}

[a] Proposed *D*-locus antigen

to confirm this association. In the 1972 International Histocompatibility Workshop, more than 1500 patients were tested and it was shown that in the patient group there were slight increases in frequencies of A1, B5, B8, and B18.

Among the other malignant diseases, retinoblastoma has the highest relative risk with the HLA specificity Bw35. Individuals with B12 may have some resistance to this disease, as seen by the low relative risk of 0.3. Nasopharyngeal cancer, a disease that is rare outside of the Chinese population, was reported to have an association with the recently described HLA-B specificity Hs (or Sin-2) (52). Multiple myeloma, Burkitt's lymphoma, malignant melanoma, mammary carcinoma, and trophoblastic neoplasia did not show significant associations with any HLA specificities.

In childhood acute lymphocytic leukemia, the A2 antigen is increased if a patient population is tested at the time of diagnosis and is even more substantially increased among long-term survivors. These data suggest that individuals with the A2 antigen are at a questionably higher risk of developing the disease, but that patients with A2 have a better prognosis than those without A2 (53).

Chronic lymphocytic leukemia has not shown any association with specific HLA antigens; however, in family studies susceptibility to this disease seems to be linked to the *HLA* locus (54).

Miscellaneous (Table 10)

Sjögren's syndrome has an association with B8, and systemic lupus erythematosus has conflicting reports of an association with Bw15.

No significant associations were observed for sarcoidosis, chronic glomerulonephritis, polycystic kidney disease, acquired hemolytic anemia and idiopathic thrombocytopenic purpura. In family studies, hay fever (ragweed pollen allergy) was reported to be linked to the *HLA* locus; however, population studies fail to show any significant associations (55). Also linked to the *HLA* locus are genes regulating quantitative levels of several of the components of the complement system (56, 57).

Table 10 Miscellaneous

Disease	HLA antigen	% Antigen frequency patient	% Antigen frequency control	Relative risk	p Value
Sjögren's syndrome	B8	51	24	3.3	0.19×10^{-6}
Congenital heart malformation	A2	80	44	4.9	0.13×10^{-5}
Acute anterior uveitis	B27	57	8	15.4	1.0×10^{-10}
Vogt-Koyanagi-Harada's disease	LD AH[a]	67	16	10.5	0.3×10^{-6}
Open-angle glaucoma	Bw35	47	23	3.0	0.48×10^{-3}

[a] Proposed *D*-locus antigen

Among congenital deformities, congenital heart malformation was associated with A2, but spina bifida and cleft lip had no HLA associations. Despite the lack of an association with any specific antigen, in one family study spina bifida did appear to be linked to the *HLA* locus (58).

Acute anterior uveitis has a strong association with B27, possibly because this disorder is so frequently found as a complication of the B27-associated forms of arthritis. In contrast, patients with aortic insufficiency, another frequent complication of AS, do not have an increased B27 frequency (59). Diffuse bilateral granulomatous uveitis associated with poliosis, vitiligo, alopecia, dysacusis, and vertigo is called Vogt-Koyanagi-Harada's disease. This disease is extremely rare in Caucasians, but fairly common in Japanese. A recent report from Japan associated this disease with Bw22J, a Japanese-specific W22-related *HLA-B* locus antigen, and even more strongly with a Japanese-specific new *D*-locus antigen LD AH (60). Finally, open-angle glaucoma has been shown to have an association with Bw35 (61).

Clinical Implications

In the preceding subheadings of this section, we have done little more than list the specific associations between HLA-antigenic specificities and particular diseases. The detailed analysis of many of these associations is just beginning, and the reader interested in a particular disease category is urged to read the source references for further details. Thus, several of the references describing the association between MS and HLA-Dw2 indicate that Dw2-positive patients with MS have a more severe, unremitting course than those patients who are Dw2-negative. Further analysis of this finding may permit subclassification of patients with MS.

In psoriasis vulgaris there is a well-known biphasic curve seen when plotting the age of onset of the disease, with peaks at about 15 years of age and again in the 30s and 40s. However, those patients who are HLA-B13- and B17-positive fall almost exclusively under the first peak, with the peak age at onset in the mid-teens. Analysis of families in which both psoriasis and B13 or B17 are segregating indicates linkage between the HLA complex and susceptibility to psoriasis. Other genetic factors not linked to HLA also have a major effect on susceptibility, since in these kindreds numerous individuals can be identified who have the disease but are HLA-B13- or B17-negative. These findings indicate that multiple genetic factors determine susceptibility to psoriasis, and that there may well be several types. This conclusion is further supported by the finding, noted earlier, that other types of psoriasis, such as pustular psoriasis, are not associated with a particular HLA type.

Analysis of a number of "autoimmune" diseases, including idiopathic Addison's disease, chronic active hepatitis, MG, Sjögren's syndrome and possibly celiac disease, dermatitis herpetiformis, Graves's disease, and insulin-dependent, juvenile-onset diabetes mellitus are all associated with HLA-B8, and many of them show an even stronger association with HLA-Dw3. It is not yet clear whether each of these diseases is associated with a particular allele of different genes, all of which are in linkage disequilibrium with HLA-B8 and HLA-Dw3, or whether all of these diseases are associated with the same allele of the same gene and there is a common initiating pathogenetic event in this very diverse list of diseases. These are but a few of the clinical correlations and questions that have been raised by an analysis of HLA and disease associations, that promise to provide important new insights into diagnostic and prognostic classification and to studies of etiology and pathogenesis.

At present, the mechanism underlying the association between HLA and specific diseases has not yet been worked out for any of the known disease associations. A knowledge of the mechanism of HLA and disease association is obviously of central importance in understanding the way in which the HLA system regulates disease susceptibility, as well as in understanding the pathogenesis of each of the associated diseases. In the absence of specific information, we are forced to deal with theoretical possibilities, and these are taken up in the next section.

POSSIBLE MECHANISMS OF HLA AND DISEASE ASSOCIATIONS

In exploring the possible mechanisms that might explain the numerous associations between the HLA system and disease, we are of necessity forced to rely partly on guesswork, and partly on our fragmentary knowledge of the known functions of genes within the major histocompatibility system in both animal and human studies. Since the function of the *H-2K* and *D*, and *HLA-A*, *B*, and *C* loci are as yet unknown, and since it is not certain that all of the functions of genes in the *I* and *S* regions of the mouse or in the *HLA-D* and complement-determining genes of man are known, it is unlikely that any theoretical list of possible mechanisms will include all the actual mechanisms yet to be discovered. On the contrary, it is quite probable that a detailed analysis of the mechanism of HLA and disease associations will lead to the discovery of new functions of the major histocompatibility system. The following is a list of the theoretical mechanisms for HLA and disease association:

1. HLA antigens may function as cell-surface receptors for viruses or for toxins.
2. HLA antigens may be incorporated into the protein coat of viruses budding from the cell surface. If this were the case, viruses from one individual would more readily infect another individual sharing the same major histocompatibility antigen.

3. A virus or toxin might share antigenic configurations with histocompatibility antigens, thus rendering the host with a particular allele of a histocompatibility antigen tolerant to those determinants on the virus or toxin.
4. *H-2K* and *D*, and *HLA-A* and *B* are known to be intimately involved in the specificity of cytotoxic T cells sensitized to viruses, minor transplantation antigens, or chemical haptens conjugated to the cell surface. It is theoretically possible that particular H-2K or D or HLA-A or B antigens might lead to more or less effective T-cell–mediated cytotoxicity against a particular virus, thus affecting the immune response to that virus.
5. Associations between HLA and disease might reflect the effect of a linked immune response (*Ir*) or immune suppression (*Is*) gene that is in linkage disequilibrium with the *HLA-A, B,* or *D* allele associated with the disease in question.
6. Abnormalities in structural genes for complement components may affect disease susceptibility. There is a very strong association between the haplotype A10-Bw18-Dw2 and C2 deficiency. While many individuals who are C2-deficient may be normal, there appears to be a high frequency of a disease resembling systemic lupus erythematosus in these individuals (62). There may also be a higher incidence of lupus-like disease in heterozygotes. While the mechanism is unknown, it seems clear from these findings that complement deficiency may predispose to certain types of "autoimmune" disease.
7. The associations between HLA and disease may be due to other linked genes, in linkage disequilibrium with particular *HLA-A, B,* or *D* alleles, but unrelated to these genes. For example, it has been shown in mice that cyclic AMP levels and testosterone levels are correlated with *H-2* type (63, 64). It is also possible that the function of a wide variety of cell-surface receptor molecules for hormones, neurotransmitters, etc is affected by different alleles of the *HLA-A, B,* or *D* regions.

Finally, it is possible that the major histocompatibility antigens have a fundamental effect on cell-cell recognition and interaction in both early development and the maintenance of normal cell interactions in the adult. Different alleles of these cell-surface proteins might permit normal development and cell-cell interaction, but, in a way as yet unsuspected, predispose to disease upon exposure to particular viruses, toxins, or environmental agents such as gluten, the factor in wheat initiating the intestinal damage in celiac disease.

With this background of theoretical possibilities, we can now examine the general characteristics of HLA and disease associations. First, there are virtually no instances in which the HLA type associated with the disease is found in all of the patients with that disease, or confers a 100% probability of contracting the disease. On the contrary, almost all of the associations are ones in which a particular allele of one of the genes in the HLA system is found in a markedly increased incidence in the patient population, but not in 100% of the patients, and the majority of individuals in the population bearing the specified allele of the *HLA* gene being studied are normal and do not have the disease. In many cases, this can be shown in family studies to be due to the fact that other genetic factors also predispose to the disease in question. In some situations (e.g. ankylosing spondylitis, where 95%

of patients have HLA-B7), it is possible that the actual disease-susceptibility gene is found in only 20% of B27-positive haplotypes, and all of these individuals develop some manifestations of the disease. Controlled family studies will be required to resolve this point.

The second major generalization is that almost all of the HLA and disease associations are for susceptibility rather than resistance, and the susceptibility effect is dominant, since most individuals are heterozygous at all of the known genetic loci in the HLA system. This generalization turns out to be of little value in analysis, since all of the genes in the HLA system appear to be codominantly expressed (i.e. the heterozygote expresses both alleles of the gene) and since almost all of the theoretical mechanisms described above, with the exception of complement deficiency, would be dominant in the heterozygote.

The third generalization is that many, if not most, of the HLA and disease associations appear to be due to genes in the *HLA-D* region, or to genes that show the strongest linkage disequilibrium with genes in the *HLA-D* region. Table 11 classifies HLA and disease associations as being due to associations with a *B* locus; with the *B* locus, but reflecting a stronger association with a *D*-locus allele; or as an association with the *D* locus only. Most of these associations either show a stronger association with an *HLA-D* allele or the patient populations have not yet been typed for their *HLA-D* type. This table reflects what has developed as a general trend over the past several years. Frequently, initial studies have reported a disease as being associated with an *HLA-A*- or *B*-locus antigen. Further study has shown that the association is stronger with the *HLA-B* antigen, and where studies have

Table 11 HLA and disease associations

Disease	B-Locus association	B- and D- association	D-Locus association only
Ankylosing spondylitis	X		
Reiter's disease	X		
Rheumatoid arthritis			X
Diabetes mellitus		X	
Addison's disease		X	
Hyperthyroidism		X	
Multiple sclerosis		X	
Myasthenia gravis		X	
Gluten-sensitive enteropathy	X	?	
Dermatitis herpetiformis	X	?	
Psoriasis	X	?	
Chronic active hepatitis	X	?	
Sjögren's syndrome	X	?	

been carried out using homozygous MLC-typing cells, the association with *HLA-B* has been found to be a function of a stronger association with an *HLA-D* allele in linkage disequilibrium with the *HLA-B* allele. Before turning to the significance of this trend, it should be noted that ankylosing spondylitis, Reiter's disease, and hemochromatosis appear to be exceptions to this generalization. Controlled family studies will be required to resolve the reason for these exceptions, but if they are due to stronger associations with *HLA-A* or *HLA-B* antigens, it would seem clear that there must be several different mechanisms underlying HLA and disease associations.

Since most HLA and disease associations appear to be due to the effect of genes either in the *HLA-D* region, or showing strongest linkage disequilibrium with this region, we must, for the moment, assume that an understanding of the function of genes in the *HLA-D* region will permit us to understand the mechanism of this type of association. Unfortunately, our knowledge of the function of the *HLA-D* gene(s) is fragmentary. At present, it is clear that these genes determine the structure of lymphocyte cell-surface antigens, eliciting a mixed-lymphocyte-culture reaction. Preliminary evidence indicates that this genetic region also determines the structure of lymphocyte cell-surface antigens analogous to the murine Ia antigens. It is probable that these antigens are the same ones that elicit the mixed-lymphocyte-culture reaction.

Beyond this, we are forced to rely on analogy with the murine model, where genes determining the structure of the Ia antigens are genetically inseparable from genes determining immune responsiveness (*Ir*) and immune suppression (*Is*), as well as cellular interaction between T lymphocytes and B lymphocytes, and T lymphocytes and macrophages. The evidence that Ia antigens are intimately involved in antigen-specific, T-cell–derived, "receptor" molecules has been cited earlier. By extrapolation, we can then postulate that many of the HLA and disease associations may be due to the effects of specific *Ir* or *Is* genes that result in hypernormal response to environmental agents, or conversely, to suppression of the normal response to environmental agents. Thus, suppression of the normal cell-mediated immune response to a virus, toxin, or environmental antigen might predispose to disease, even though superficial measurement of antibody response to the virus in question reveals no significant differences.

Although all of the available evidence indicates that HLA-linked *Ir* and *Is* genes are major factors in HLA and disease associations, it must be emphasized that *Ir* genes have not yet been conclusively demonstrated in man. In addition, the several HLA-linked complement genes have not yet been precisely mapped. It is quite possible that the genetic organization of the HLA system shows considerable differences from the H-2 system, a possibility underlined by the existence of the *HLA-C* locus, which does not yet have a murine counterpart. The different recombination rates between *HLA-A* and *B*, and *H-2K* and *D*, and the different position of the *HLA-D* gene(s) and the *I*-region genes emphasizes the possibility that the genetic organization of the HLA system may show important differences from the H-2 system, despite the very close overall analogy in the structure and molecular weight of the gene products of both of the systems. Noting these cautions, the actual and potential analogies are still impressive. Thus, in the mouse it has been demonstrated (see Section 2) that two genes in the *I* region are required for immune responsiveness to some antigens. This finding offers a potential explanation for the marked linkage disequilibrium seen in the HLA system and postulated to account for the association between particular *HLA-B* or *D* alleles and disease. Thus, a particular *HLA* haplotype might be selected for because it carried two genes required for immune response (or conversely, suppression of the immune response) to a particular foreign antigen.

Study of this possibility will require identification of the numerous genes postulated to exist in the *HLA-D* region of the HLA system. It is probable that the *HLA-D* region is as complex as the *I* region in the mouse, and that particular combinations of alleles of different genes in the *HLA-D* region have marked selective survival advantage. While these haplotype combinations may have conferred a survival advantage during the course of evolution, it is also possible that particular alleles of one or more genes within the *HLA-D* region predispose to a particular disease. (In this respect, it is important to note that most of the HLA-associated diseases do not affect the individual until after the peak reproductive period, and would thus not function as semilethals.)

CONCLUSIONS AND PROSPECTS

It is now clear that the HLA system plays a major role in determining susceptibility to a wide variety of diseases. Many, but not all of these HLA-disease associations are due to genes in the *HLA-D* region, the human analog of the murine *I* region. It seems clear that many of these disease associations are due to the effect of immune response (*Ir*) or immune suppression (*Is*) genes determining hyper- or hyponormal responsiveness to viruses, toxins, or environmental antigens. It seems equally clear that this extraordinarily complex genetic system has other functions, and that derangements in some of these other functions underlie some of the observed HLA-disease associations.

Our knowledge of the genetic organization of the HLA system is presently very incomplete, particularly with respect to genes other than those determining the serologically detectable antigens (*HLA-A, B, C*). Because of the importance of the *HLA-D* region in disease association, and because of the complexity of the analogous *I* region in the mouse, the next major step in analyzing the function of the HLA system and its disease associations will involve development of techniques for detecting particular alleles of the multiple genes in the *HLA-D* region, which presumably code for cell-surface proteins analogous to murine Ia antigens. The demonstration that some pregnancy sera contain antibodies that react primarily with B lymphocytes, and which also identify a set of cell-surface alloantigens on B lymphocytes of approximate molecular weight 33,000, indicates that specific sera are available. Analysis of this alloantigenic system may permit subdivision and identification of alleles at closely linked loci in a manner analogous to the analysis of the *HLA-A, B,* and *C* loci.

The ability to detect different alleles of the multiple genes that almost certainly exist in the *HLA-D* region offers great promise in further analyzing and understanding associations between the HLA system and disease. First, correlation of these postulated genes with immune responsiveness to a wide variety of antigens should permit conclusive demonstration and mapping of human *Ir* and *Is* genes. Second, it is likely that identification of multiple genes within the *HLA-D* region will result in demonstration of extremely strong associations between particular alleles of genes in this region and particular diseases. Thus, if adequate reagents were available, it might be possible to demonstrate that all or nearly all patients with a particular autoimmune disease such as Sjögren's syndrome or chronic active hepatitis carried a particular allele of one or two genes in the *HLA-D* region, whereas current methods of *HLA-D* locus typing (MLC typing) can detect only those patients who share all genes in the *HLA-D* region.

If such strong associations can be detected for those diseases known to be associated with the HLA system, it is also likely that new associations between particular genes in the *HLA-D* region and other diseases will be discovered. The list of candidate diseases is large (e.g. polyarteritis nodosa, scleroderma, dermatomyositis, Guillain-Barré syndrome, Wegner's granulomatosis, polymyalgia rheumatica, etc). If new and stronger associations are discovered with *HLA-D* region genes, then HLA typing, and in particular *HLA-D* typing, may well be extremely important diagnostic and prognostic tools. For those diseases showing strong associations with *HLA-D* region genes, abnormalities in immune responsiveness, either hyper- or hyponormal, may well be major predisposing factors, and it will be important to identify the environmental agent—virus, toxin, or environmental antigen—responsible for initiating the pathogenic process.

Finally, we hope that studies of the function of the *H-2K* and *H-2D*, and *HLA-A*, *B*, and *C* gene products, as well as a study of their structure, will lead to an understanding of the function of these gene products, their polymorphism, and their relationship to other cell-surface molecules, such as receptor molecules. This in turn should permit an understanding of the way in which different alleles of these genes might predispose to diseases such as ankylosing spondylitis and hemochromatosis. The recent findings that diseases such as hemochromatosis and primary open-angle glaucoma show an association with genes in the HLA system underlines the extraordinarily broad effects of this system, and emphasizes the probability that the major histocompatibility system plays a major role in the function of cell-surface molecules and in cellular interaction. Unraveling all of these complex functions promises to make major contributions to our understanding of the pathogenesis of a wide variety of diseases.

ACKNOWLEDGMENTS

This work was supported in part by National Institutes of Health Grants AI 07757 and AI 11313.

Literature Cited

1. Lilly, F., Boyse, E. A., Old, L. J. 1964. Genetic basis of susceptibility to viral leukaemogenesis. *Lancet* ii:1207–9
2. Amiel, J. L. 1967. In *Histocompatibility Testing,* ed. E. S. Curtoni, P. L. Mattius, R. M. Tosi, pp. 79–81. Copenhagen: Munksgaard. 458 pp.
3. McDevitt, H. O., Tyan, M. L. 1968. Genetic control of the antibody response in inbred mice: Transfer of response by spleen cells and linkage to the major histocompatibility (H-2) locus. *J. Exp. Med.* 128:1–11
4. McDevitt, H. O., Benacerraf, B. 1969. Genetic control of specific immune responses. *Adv. Immunol.* 11:31–74
5. Benacerraf, B., McDevitt, H. O. 1972. The histocompatibility-linked immune response genes. *Science* 175:273–79
6. Katz, D. H., Benacerraf, B. 1976. *The Role of Products of the Histocompatibility Gene Complex in Immune Responses.* New York: Academic. 780 pp.
7. Shreffler, D. C., David, C. S. 1975. The *H-2* major histocompatibility complex and the *I* immune response region: Genetic variation, function, and organization. *Adv. Immunol.* 20:125
8. Zinkernagel, R. M., Doherty, P. C. 1976. See Ref. 6, pp. 203–12
9. McDevitt, H. O., Deak, B. D., Shreffler, D. C., Klein, J., Stimpfling, J. H., Snell, G. D. 1972. Genetic control of the immune response: Mapping of the *Ir-1* locus. *J. Exp. Med.* 135:1259–78
10. Benacerraf, B. 1976. See Ref. 6, pp. 225–48
11. Taussig, M. J., Munro, A. J., Luzzati, A. L. 1976. See Ref. 6, pp. 541–52
12. Tada, T., Taniguchi, M. 1976. See Ref. 6, pp. 506–12
13. Murphy, D. B., Okumura, K., Herzenberg, L. A., Herzenberg, L. A., McDevitt, H. O. 1976. Selective expression of separate *I* region loci in functionally different lymphocyte subpopulations. *J. Exp. Med.* In press
14. McDevitt, H. O. 1976. The evolution of genes in the major histocompatibility complex. *Fed. Proc.* 35:2168–73
15. Dausset, J. 1954. Leuco-agglutinins. IV. Leuco-agglutinins and blood transfusion. *Vox Sang.* 4:190–98
16. Payne, R. 1957. Leukocyte agglutinins in human sera. *Arch. Intern. Med.* 99:587–606
17. van Rood, J. J., van Leeuwen, A., Schippers, A., Vooys, W. H., Frederiks, H., Balner, H., Eernisse, J. G. 1965. in *Histocompatibility Testing,* ed. D. B. Amos, J. J. van Rood, pp. 37–50. Copenhagen: Munksgaard. 288 pp.
18. Payne, R., Tripp, M., Wigle, J., Bodmer, W., Bodmer, J. 1964. A new leukocyte isoantigen system in man. *Cold Spring Harbor. Symp. Quant. Biol.* 29:285–95
19. Dausset, J., Ivanyi, P., Ivanyi, D. 1965. See Ref. 17, pp. 51–62
20. Terasaki, P. I., McClelland, J. D. 1964. Microdroplet assay of human serum cytotoxins. *Nature* 204:998–1000
21. Kissmeyer-Nielsen, F., Thorsby, E. 1970. Human transplantation antigens. *Transplant. Rev.* 4:1–176
22. Thorsby, E. 1974. The human major histocompatibility system. *Transplant. Rev.* 18:51-129
23. Svejgaard, A., Hauge, M., Jersild, C., Platz, P., Ryder, L. P., Staub-Nielsen, L., Thomsen, M. 1975. The HLA system: An introductory survey. *Monogr. Hum. Genet.* 7:3–102
24. Lammn, L. U., Friedrich, U., Petersen, G. B., Jorgenson, J., Nielsen, J., Therkelsen, A. M., Kissmeyer-Nielsen, F. 1974. Assignment of the major histocompatibility complex to chromosome No. 6 in a family with pericentric inversion. *Hum. Hered.* 24:243–84
25. Kissmeyer-Nielsen, F., Svejgaard, A., Hauge, M. 1968. Genetics of the human HL-A transplantation system. *Nature* 219:1116–19
26. Low, B., Messeter, L., Mansoon, S., Lindholm, T. 1974. Crossing-over between the SD-2 (four) and SD-3 (AJ) loci of the human major histocompatibility chromosomal region. *Tissue Antigens* 4:405
27. Dupont, B., Staub-Nielsen, L., Svejgaard, A. 1971. Relative importance of four and LA loci in determining mixed-lymphocyte reaction. *Lancet* ii:1336–40
28. Bodmer, J. G., Colombani, J., Rocques, P., Degos, L., Bodmer, W. F., Dausset, J., Piazza, A. 1972. In *Histocompatibility Testing,* ed. J. Dausset, J. Colombani, pp. 621–67. Copenhagen: Munksgaard. 778 pp.
29. Bodmer, J. 1975. In *Histocompatibility Testing.* ed. F. Kissmeyer-Nielsen, pp. 21–99. Copenhagen: Munksgaard. 1035 pp.
30. Payne, R., Radvany, R., Grumet, F. C. 1975. A new second locus HL-A antigen in linkage disequilibrium with HL-

A2 in Cantonese Chinese. *Tissue Antigens* 5:69–71

31. Strominger, J. L., Chess, L., Herrmann, H. C., Humphreys, R. E., Malenka, D., Mann, D., McCune, J. M., Parham, R., Robb, R., Springer, T. A. Terhorst, C. 1975. See Ref. 29, pp. 719–30
32. Goodfellow, P., Jones, E., van Heyningen, V., Solomon, E., Bobrow, M., Miggiano, V., Bodmer, W. F. 1975. The β-2 microglobulin gene is on chromosome #15 and not in the HL-A region. *Nature* 254:267–69
33. van den Tweel, J. G., Blusse van Oud Alblas, A., Keuning, J. J., Goulmy, E., Termijtelen, A., Bach, M. S., van Rood, J. J. 1973. Typing for MLC (LD): I. Lymphocytes from cousin-marriage offspring as typing cells. *Transplant. Proc.* 5:1535–49
34. Thorsby, E., Piazza, A. 1975. See Ref. 29, pp. 414–58
35. Haverkorn, M. J., Hofman, B., Masurel, N., van Rood, J. J. 1975. HL-A linked genetic control of immune response in man. *Transplant. Rev.* 22: 120–24
36. McDevitt, H. O., Delovitch, T. L., Press, J. L., Murphy, D. B. 1976. Genetic and functional analysis of the Ia antigens: Their possible role in regulating the immune response. *Transplant. Rev.* 30:197–235
37. Wernet, P. 1976. Human Ia-type alloantigens: Methods of detection, aspects of chemistry and biology, markers for disease states. *Transplant. Rev.* 30:270–98
38. Walford, R. L., Gosset, T., Troup, G. M., Gatti, R. M., Mittal, K. K., Robbins, A., Ferrara, G. B., Zeller, E. 1976. The Merritt alloantigenic system of human B-lymphocytes. Evidence for 13 possible factors including one six member segregant series. *J. Immunol.* 116:1704–10
39. Ting, A., Mickey, M. R., Terasaki, P. I. 1976. B-lymphocyte alloantigens in Caucasians. *J. Exp. Med.* 143:981–86
40. Jersild, C., Fog, T., Hansen, G. S., Thomsen, M., Svejgaard, A., Dupont, B. 1973. Histocompatibility determinants in multiple sclerosis with special reference to clinical course. *Lancet* ii:1221–25
41. Terasaki, P. I., Park, M. S., Opelz, G., Ting, A. 1976. Multiple sclerosis and high incidence of a B lymphocyte antigen. *Science.* In press
42. Ryder, L. P., Svejgaard, A. 1976. In *Report from the HLA and Disease Registry of Copenhagen 1976,* pp. 1–34 Copenhagen: Ryder & Svejgaard. 34 pp.
43. Svejgaard, A., Platz, P., Ryder, L. P., Staub-Nielsen, L., Thomsen, M. 1975. HLA and disease associations—A survey. *Transplant. Rev.* 22:3–44
44. Calin, A., Fries, J. F. 1975. Striking prevalence of ankylosing spondylitis in "Healthy" W27 positive males and females. A controlled study. *N. Engl. J. Med.* 293:835–39
45. Sairanen, E., Tillikinan, A. 1975. HL-A27 in Reiter's Disease following shigellosis. *Scand. J. Rheumatol.* 4: 30–41
46. Calin, A., Fries, J. F. 1976. An "experimental" epidemic of Reiter's syndrome revisited. Follow-up evidence on genetic and environmental factors. *Ann. Intern. Med.* 84:564–66
47. Wagner, L. P., Fessel, W. J. 1975. HL-A27 (W27) absent in gonococcal arthritis. *Lancet* i:1094–95
48. Morris, R., Metzger, A. L., Bluestone, R., Terasaki, P. I. 1974. HL-A W27—A clue to the diagnosis and pathogenesis of Reiter's syndrome. *N.Engl. J. Med.* 290:554–56
49. Stastny, P. 1975. See Ref. 29, pp. 797–804
50. McMichael, A. J., Sasazuki, T., McDevitt, H. O., Payne, R. O. Manuscript in preparation
51. Mann, D. L., Katz, S. J., Nelson, D. L., Abelson, L. D., Strober, W. 1976. Specific B cell antigens associated with gluten-sensitive enteropathy and dermatitis herpetiformis. *Lancet* i:110–11
52. Simons, N. J., Chan, S. H., Ho, J. H. C., Chan, J. C. W., Day, N. E., de The, G. B. 1975. See Ref. 29, pp. 809–12
53. Rogentine, G. N., Trapani, R. J., Yankee, R. A., Henderson, E. S. 1973. HL-A antigens and acute lymphocytic leukemia: The nature of the HL-A2 association. *Tissue Antigens* 3:470–76
54. Delmas-Marsalet, Y., Hors, J., Colombani, J., Dausset, J. 1974. Study of HL-A genotypes of a case of familial chronic lymphocytic leukemia (CLL). *Tissue Antigens* 4:441–45
55. Levine, B. B., Stember, R. H., Fotino, M. 1972. Ragweed hayfever: Genetic control and linkage to HLA haplotypes. *Science* 178:1201–3
56. Fu, S. M., Kunkel, H. G., Brusman, H. P., Allen, F. H., Fotino, M. 1975. Evidence for linkage between HL-A histocompatibility genes and those involved in the synthesis of the second

component of complement. *J. Exp. Med.* 140:1108–12

57. Rittner, C., Hauptman, G., Grosse-Wilde, H., Grosshaus, E., Tongio, M. M., Mayer, S. 1975. See Ref. 29, pp. 945–54
58. Amos, D. B., Ruderman, R., Mendell, N. R., Johnson, A. H. 1975. Linkage between HL-A and spinal development. *Transplant. Proc.* 7:93–95
59. Calin, A., Fries, J. F., Stinson, E. B., Payne, R. 1976. Normal frequency of HL-A B27 in aortic insufficiency. *N. Engl. J. Med.* 293:397–99
60. Tagawa, Y., Sugiura, S., Yakura, H., Wakisaka, A., Aizawa, M., Itakura, K. 1976. In *HLA and Disease* ed. J. Dausset, A. Svejgaard, p. 263 Paris: Inst. Natl. Sante Rech. Med. 333 pp.
61. Aviner, Z ., Henley, W. L., Fotino, M., Leopold, I. H. 1976. Histocompatibility (HLA) antigens and primary open angle glaucoma. *Tissue Antigens* 7:193–200
62. Stem, R., Fu, S. M., Fotino, M., Agnello, V., Kunkel, H. G. 1976. Hereditary C2 deficiency associations with skin lesions of systemic lupus erythematosus. *Arthritis Rheum.* 19: 517–19
63. Meruelo, D., Edidin, M. 1975. Association of mouse liver adenosine 3':5'-cyclic monophosphate (cyclic AMP) levels with histocompatibility-2 genotype. *Proc. Natl. Acad. Sci. USA* 72: 2644–48
64. Ivanyi, P., Hampl, R., Starka, L., Mickova, M. 1972. Genetic association between H-2 gene and testosterone metabolism in mice. *Nature New Biol.* 283:280–81

Ann. Rev. Med. 1977. 28:453-65
Copyright © 1977 by Annual Reviews Inc. All rights reserved

STRUCTURE AND FUNCTION OF ALPHA-FETOPROTEIN

❖7254

Thomas B. Tomasi, Jr., M.D., Ph.D.
Department of Immunology, Mayo Medical School, Rochester, Minnesota 55901

INTRODUCTION

Alpha-fetoprotein (AFP) is one of several embryo-specific proteins and is a dominant serum protein as early in human embryonic life as one month, when albumin and transferrin are present in relatively small amounts (1). It is first synthesized in the human by the yolk sac and liver (1–2 months) and subsequently predominantly in the liver (2). In other species, such as the mouse, rat, and chicken, synthesis by the yolk sac continues until birth. A small amount of AFP is produced by the GI tract of the human conceptus. In sharks a protein, perhaps analogous to AFP, has been described that is produced in the yolk sac and liver and also in large amounts in the GI tract (3). Thus, a homologue of AFP may exist as far back in evolution as 400 million years, at a time when the mammals and elasmobranchs diverged.

The function of AFP during development and subsequently in the adult is largely unknown. Interest in this protein has centered primarily on the clinical observations (see below) that it may reappear in the serum in elevated amounts in adult life in association with normal restorative processes (i.e. liver regeneration) and with malignant growth. Elevations of AFP are a well-known diagnostic aid in hepatomas and teratoblastomas, but in addition, more recent data suggest that a significant percentage of other tumors, particularly of primitive gut origin, also have elevated levels (4). It is thought that the reappearance of this fetal protein results from the decrease of a repressor, which normally appears during the end of embryonic life.

STRUCTURE

Mouse AFP is a glycoprotein (5% carbohydrate) having a single polypeptide chain (single N-terminal residue) with a molecular weight of 70,000 by SDS gel electrophoresis, and about 67,000 by ultracentrifugation (T. Tomasi, K. Oles, unpublished data). The molecular weights and amino acid compositions of human, dog, rabbit, cat, rat, and mouse AFP are similar, although species differences are found in compositions and peptide mapping (5). At least two types of rat AFP (6) and three types of human AFP (7) have been recognized by gel electrophoresis. The basis of these forms and whether they represent synthetic or postsynthetic differences is unknown. Although AFPs with differing numbers of sialic acid residues have been recognized, the forms referred to above are not due to differences in sialic acid content. An interesting study (8) indicates that AFP and albumin show significant amino acid homologies, suggesting the possibility that AFP could be an embryonic analogue of albumin, similar to the fetal-adult hemoglobin system.

Immunization of an animal with its own AFP does not lead to antibody formation because of tolerance, probably of the low-zone type, maintained by T cells. Tolerance can be broken by altering the AFP (i.e. addition of haptenic groups) or by immunizing with a cross-reacting AFP from another species (9). For example, rabbits immunized with human AFP react both with human and with their own AFP (5), although maintaining some specificity for the human protein. These antibodies eliminate the animal's normal serum AFP, demonstrating that they are indeed autoantibodies (9).

SYNTHESIS

AFP production has been studied in several in vitro systems, including cell lines derived from rat hepatomas and isolated fetal hepatocytes. Using rat hepatoma lines, it was shown that some clones produce high levels of AFP and low levels of albumin, and others the reverse (10). Moreover, albumin and AFP are produced during different parts of the cell cycle. AFP is synthesized during G1 and S phases, while albumin is produced from mid-S through G2, the exact phases in the cell cycle varying somewhat with different clones. By immunofluorescence, it has been determined that AFP and albumin are probably synthesized in different cells and hepatomas consist of mixtures of cells, some producing AFP, others albumin, and some nonproducing cells. If a high proportion of nonproducing cells is present, the serum levels of AFP may be low (10). The capacity of a given hepatoma to produce AFP is also related to its chromosome composition and growth rate (11), as well as to the phase of the cell cycle. In general, fast-growing aneuploid tumors produce more AFP and have higher serum AFP levels, although exceptions occur. Studies on monolayers of fetal hepatocytes have suggested that production of AFP and the heme-binding protein hemopexin are related to hepatocyte proliferation, whereas albumin and haptoglobin synthesis are not (12). However, certain hepatocarcinogens (N-2 fluorenylacetamide) stimulate AFP production in the absence of a detectable increase in mitosis (13), so that DNA synthesis and mitosis are not necessary for the release of AFP. AFP levels are also influenced by hormones with glucocorticoids, thyroxine, and adrenaline, all causing a dramatic decrease in blood levels of AFP in newborn mice (14). It is postulated that AFP is produced by an "intermediate" cell, lying ontogenically between a stem cell and the mature hepatocyte, and that glucocorticoids block the proliferation of the AFP-producing cells.

A key question concerns the significance of the coupling of AFP synthesis to the cell cycle. A possibility to be seriously considered is that AFP in some way controls cell growth, although exogenously added AFP does not affect the rate of cell growth of fetal hepatocytes (15). It has been hypothesized (16) that intracellular AFP may affect cell growth by virtue of its ability to bind estrogens. Rat and mouse AFP have a high binding affinity for certain estrogens (17, 18). Since estrogens stimulate very low-density lipoprotein (VLDL) output by the liver and VLDL suppresses the initiation of DNA synthesis by fetal hepatocytes, it has been speculated that AFP, by binding estrogens, lowers VLDL and therefore DNA synthesis increases (16). We have noted (19) that during a mixed-lymphocyte reaction, levels of total protein synthesis and membrane AFP occur in a cyclic fashion, with a decrease in protein synthesis being preceded by a rise in membrane AFP, suggesting that AFP may act as a negative-feedback signal for protein synthesis. If these initial experiments are verified, then AFP could play a role in regulation of cell growth and synthetic activities in cells other than hepatocytes, i.e. lymphocytes and possibly other cells. A recent report (20) suggests that AFP may be identical to the cytosol receptor for estrogens in the rat uterus. Since this receptor plays a key role in controlling RNA and protein synthesis in estrogen-sensitive tissues (21), the identity of AFP with the steroid receptor would indicate an important function of AFP in controlling cell metabolism. At the moment, however, it is difficult to reconcile the various observations mentioned above, and they should be considered as interesting speculations worthy of further experimentation.

METABOLISM

Each species has a characteristic pattern of changes in the concentration of AFP during development. As shown in Figure 1, AFP reaches a maximum level in human fetal serum around the third month of gestation (3.0 mg/ml), then drops almost linearly as birth approaches (22). Although the concentration of AFP in fetal serum falls after the first trimester, the total quantity of AFP synthesized increases until 20 weeks and then remains constant until about 32 weeks. Thus the decrease in serum concentration is not due to diminished synthesis, but rather to the fact that the weight of the fetus and thus the volume in which AFP is distributed is increasing even more rapidly than synthesis (22). Following birth, AFP levels decline very rapidly (half-life of human AFP is 3.5–4.0 days) during the first 1–2 weeks of life, then more slowly until it reaches normal adult levels by 1–2 years of life (23). In the rat and mouse, the levels in the fetal serum fall much less sharply in later pregnancy and are significantly higher (8–10 mg/ml) than in the human, due mainly to continued synthesis by the yolk sac. AFP synthesis in these species falls rapidly at birth (loss of yolk-sac synthesis), but continues at a higher level than in the human for 5–8 weeks, then stabilizes at adult levels.

Elevated serum levels of AFP (500 ng/ml) have been reported in normal pregnant women (24), which reach a maximum during the third trimester. The reason for the late rise compared with the peak of fetal concentration (12–14 weeks) is unknown. Although the placental size increases 3.3-fold from 20 to 32 weeks, the maternal plasma concentration increases ninefold over the same period (2). Thus, either there is an increase in placental permeability as pregnancy progresses, or the mother synthesizes AFP. There is evidence for the former with other proteins, although

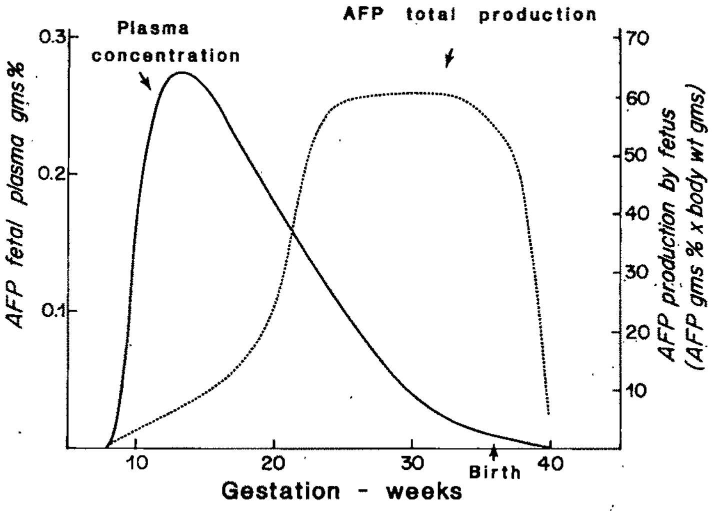


Figure 1 Human fetal plasma concentration of AFP (———) and total production of AFP measured by plasma concentration X fetal body weight (--------) at various times during human gestation. (Redrawn from Gitlin, D., Boesman, M. 1966 *J. Clin. Invest.* 45:1826.)

maternal synthesis has not been as carefully examined as it should be. Since the development of a sensitive radioimmunoassay, it is now known that AFP persists in normal adult serum in amounts that are approximately 1/10,000 of the maximal fetal concentration. In the human, the normal adult serum values are about 1–40 ng/ml, whereas certain strains of adult mice normally have 200–500 ng/ml in their serum.

IMMUNOREGULATION BY AFP

An intriguing possibility suggested by current work from several laboratories (25–32) is that AFP has immunoregulatory properties, although not all reports are in agreement with this finding (33–36). Such a function, if verified, could be important in the regulation of the immune response during the fetal and newborn period, and also have implications in other situations where AFP is significantly elevated, such as in hepatitis, tumors, and during pregnancy.

Basically, the theory that AFP has immunoregulatory properties rests on the following observations. Microgram amounts of purified AFP added to in vitro cultures of splenic or lymph-node mononuclear cells exert a suppressive effect on antibody synthesis, both in the primary and secondary responses to the T-cell–dependent antigen, sheep red blood cells. The effect is not due to cytotoxicity, and preincubation studies indicate that lymphocytes treated with AFP for eight hours and then washed are suppressed. Administration of AFP in the form of mouse amniotic fluid also inhibits the in vivo antibody response, particularly in the T-cell–dependent classes (γG and γA) (27). Cell-mediated reactions, such as mitogen-induced lymphocyte transformation, as well as the mixed-lymphocyte reaction (MLR), are also inhibited by the order of 1–10 μg/ml of AFP added to the culture. Some confusion exists regarding a recent report (37) that bovine fetuin is immunosuppressive. However, bovine fetuin and AFP are quite different proteins, and we have recently shown (38) that isolated bovine AFP is suppressive in small concentrations (1–5 μg/ml), while fetuin is stimulatory in concentrations up to 500 μg/ml.

The cell involved in binding mouse AFP is a subset of T lymphocytes (39). This is based primarily on work using the fluorescent-antibody technique showing that cell-surface binding of AFP occurs after in vitro incubation of AFP with preparations of spleen cells containing 95% T cells. Approximately one third of the T cells in these preparations have receptors for AFP, and no binding occurs to macrophage-enriched populations, nude spleen cells, spleen cells depleted of T cells by treatment with anti-Θ plus complement, or bone marrow cells. Interestingly, there is little binding to whole thymocytes and only 6% of cortisone-resistant thymocytes have receptors for AFP. This suggests the possibility that AFP surface receptors require additional maturation after peripheralization from the thymus. Splenic lymphocytes from pregnant animals also have about one third of their T cells bearing surface AFP. These cells are a special subset of light-density T cells that can be isolated on Ficoll density-gradient columns (R. Thompson, T. Tomasi, in press). AFP could act by stimulating suppressor cells or, alternatively, by inhibiting helper and amplifier cells. Preliminary data suggest that AFP effects changes in intracellular cyclic AMP (cAMP) and GMP (cGMP) levels in a direction (decreased intracellular cAMP/cGMP), suggesting that T cells are suppressed (M. Burritt, R. Keller, T. Tomasi, unpublished observations).

Interestingly, although many of the clinical disorders (discussed below) are associated with decreased T-cell numbers, human AFP does not inhibit E-rosette formation (40, 41).

ESTROGEN-BINDING PROPERTIES OF AFP

Mouse and rat AFP has a high binding affinity for estrogens ($K_d = 10^{-9}$–10^{-10}M) with estrone > estradiol > estriol. Assuming one site per molecule, it has been estimated (42) that 20–30% of fetal serum and 90% of amniotic-fluid AFP have bound estrogens. This suggests that two species of AFP may occur, with regard to their estrogen-binding activity. However, only a small percentage (< 2%) of AFP molecules in human serum has been reported to possess estrogen-binding properties. The proportion of estrogen-binding AFP in human amniotic fluid, however, is variable in our experiments and more work in this area is needed. However, using tritiated estrogens and autoradiography, labeled cells have been found in human tumors of nonhepatic as well as hepatic origin. The presence of radioactivity in fetal rat liver, yolk sac, and intestine, but not in adult animals, suggests that AFP is in large part responsible for this localization, although such studies do not provide direct evidence for AFP synthesis in these tissues. Interest in the binding properties of AFP has been heightened by the report (43) that the immunosuppressive activity of AFP can be deleted by dialysis against 0.5M KCl (pH 5.8), a procedure that does not remove estrogens from their cell receptors. Moreover, activity can be restored by the addition of an ether extract of the KCl dialysate, which contains high levels of estrogens by radioimmunoassay. However, on thin-layer chromatography of the ether extract, at least four components are seen, so the nature of the restorative substance(s) is presently uncertain, and it has not been excluded that prostaglandin or some other unidentified dialyzable factor is necessary for suppression. The sialic acid content of the AFP could also be of importance since in our work, cell-surface binding of AFP is inhibited by methyl α-D-glucopyranoside.

CLINICAL SITUATIONS WITH ELEVATED AFP

Elevations of serum AFP have been reported in (*a*) the fetus and neonatal animal; (*b*) pregnancy; (*c*) hepatitis and chronic liver diseases; (*d*) tumors, including hepatomas, teratomas, and others, particularly those of primitive gut origin; and (*e*) ataxia telangiectasia.

Fetus and Newborn

It is well established that the fetus and newborn of many mammalian species are immunologically immature in their ability to carry out certain T- and B-cell–mediated functions compared to the adult. However, there is considerable variation among species in the degree of the immune suppression and its duration following birth. There is no question, particularly from the studies by Silverstein's group (44), that under appropriate antigenic stimulation, the fetus and newborn are capable of mounting a good immunological response approximating that of the adult, although some defects are seen, depending on the antigen employed. Whether the variations noted in the immune status of neonates between species and/or with different antigens are a reflection of the levels of AFP and the differing effects of AFP on various antigens has not been studied.

An important question involves the possible survival advantages for fetal immunosuppression. Several come to mind. First, potentially antigenic substances of benefit to the fetus pass across the placenta, and were they to elicit an immune response in the fetus, they would be rapidly inactivated. For example, it has been pointed out (45) that there is a survival advantage to the passage of foreign allotypes of immunoglobulins from the mother to the fetus and that this would be seriously compromised if the offspring were able to respond to the incompatible maternal globulins, resulting in their elimination. Second, cells with a capacity for self-recognition have been found to exist in normal animals (46). In fetal life, AFP could be important in suppressing autoimmunity (maintaining tolerance) while later in development, other factors, including blocking antibodies (47), could become primary in importance. It is well known that tolerance is more readily induced in the neonatal period and AFP could regulate tolerance. Third, seeding of maternal cells into the fetus could lead to a graft-versus-host reaction. While the passage of fetal cells into the mother has been well recognized, the extent to which maternal cells reach the fetus is unclear. In any case, two alternatives would be available to the fetus for avoiding a graft-versus-host reaction—the rapid rejection of the maternal cells or the inactivation of the immunological capability of maternal cells, possibly by AFP. In the latter case, maternal cells would persist in the fetus until the concentrations of AFP fell to levels that would allow fetal cells to reject the lymphocytes. In order to avoid a graft-versus-host reaction, the fetal cells must escape suppression by AFP prior to those seeded from the mother.

Pregnancy

Nonspecific suppression of cell-mediated reactions in pregnancy has been found in a number of studies [reviewed in (48)]. These include a depression of the MLR, diminished response to mitogens, decreased tuberculin skin reactivity, decreased in vitro response to PPD, depressed skin homograft reactions, and a deficiency in cell-mediated reactions to rubella virus. A recent report (49) indicates that the percentage of peripheral blood T and B cells changes significantly during human pregnancy. From the 7th to the 20th week, an inversion in the T:B cell ratio (30% T, 70%B) occurs, which returns to normal nonpregnant levels (70%T, 30%B) by 22 weeks, where it remains throughout pregnancy. Several workers have reported that maternal plasma contains a factor suppressive for the MLR, mitogen response, and antigen-induced leukocyte migration of normal cells. It has been suggested that this suppression could be a factor in the failure of the mother to reject the fetal allograft. It has also been postulated that defective cellular immunity could be related to the increased severity and dissemination of viral infections during pregnancy, which in turn lead to congenital infections such as those with cytomegalic and rubella viruses.

Studies in our laboratory (48) have also shown significant suppression in the antibody response to sheep red blood cells in pregnant mice, particularly in the γA and γG classes. In addition, we have shown that pregnant serum is suppressive, when added in vitro, for antibody synthesis and cell-mediated reactions, and that suppression is specifically removed by affinity chromatography with an anti-AFP antiserum. This suggests that AFP could be an important factor in the immunosuppression of pregnancy. However, the interpretation of these experiments is complicated by the presence of suppressive factors in normal mouse serum, other suppressive (questionable hormonal factors) in pregnant sera, and a recent work showing significant but not complete removal of suppression of an anti-immunoglobulin affinity column.

Heterologous (sheep) antibodies against AFP injected into pregnant rabbits causes a high incidence of fetal mortality (50). It is not certain whether this results from a depletion of AFP, immune complexes, or a direct effect on some target tissue, such as the yolk sac, which synthesizes AFP

Hepatic Diseases

It is known that some patients with acute hepatitis and chronic acute liver disease (CALD) have elevated levels of AFP. It has also been reported (51) that certain patients with chronic hepatitis show diminished cell-mediated reactions. We have verified that approximately one third of patients with CALD have diminished mitogen and MLRs (52). Approximately 30% of the peripheral blood T cells isolated by a sheep red-blood-cell rosetting technique from the suppressed patients have surface AFP, while none of the nonsuppressed CALD patients and age-matched controls have AFP-positive cells. When cells from suppressed patients are cultured for seven days in vitro, surface AFP is lost and the washed cells regain normal responsiveness. A suppressive factor for cell-mediated reactions of normal cells is then present in the culture supernates. Although AFP-bearing cells showed a high correlation with suppression, they did not correlate with the presence of Hb_s Ag or other immunoserologic markers (ANA, LE, SMA), disease activity, duration, or steroid treatment. Thus the role of AFP and immune suppression in the pathogenesis of CALD remains obscure.

AFP is classically elevated in some of the patients with hepatomas, and has been a useful diagnostic tool in their detection. The serum levels of AFP in hepatomas vary widely and in some cases are normal. This is probably related to the number of AFP-producing versus nonproducing cells in the particular tumor. We have examined (R. Keller, A. Schutt, T. Tomasi) the sera of seven patients with hepatomas and found they suppressed the phytohemagglutinin (PHA) responsiveness and the MLR of normal mononuclear cells. Whether this is due to AFP has not been completely established.

The sera of patients with tyrosinemia have elevated levels of AFP, probably related to the associated hepatic disease. These children have normal cell-mediated reactions (mitogen responses and MLRs), and this is a cogent argument against an important immunoregulatory role for AFP (34). However, further studies on the nature of the AFP in these patients, its ability to suppress normal cells and the suppressibility of cells from these patients by normal AFP should be carried out before any definite conclusions are reached.

Malignancies

Malignancies other than hepatomas and teratomas, particularly neoplasms of the primitive gut origin (gastric, pancreatic, and biliary carcinoma), may have high serum levels of AFP (4). Liver metastasis need not be present for elevation of AFP to occur, and the evidence suggests that AFP may be produced locally at the tumor site. The elevations of AFP in tumors of endodermal origin may represent derepression of cells capable of synthesizing AFP during early development.

In studies of patients and their relatives from so-called cancer-prone families (53), we have found (54) a high incidence of surface AFP–positive peripheral-blood mononuclear cells in the absence of elevations of serum AFP determined by a sensitive radioimmunoassay technique. It was found that 86% of patients with overt malignancies, 63% of first-degree relatives, and 18% of second-degree relatives had AFP-bearing cells, while none of the spouses or normal age-matched controls possessed positive cells. The presence of cells with surface AFP showed a high correlation with suppression of mitogen responsiveness. Culture of the AFP-positive lymphocytes in vitro for seven days resulted in a return toward normal of the mitogen response that was associated with a loss of AFP from the cell surface. Thirty-two patients with gastric carcinoma and twelve with brain tumors from families without a striking incidence of cancer had normal mitogen responses and no AFP-staining cells. It will be of considerable interest to follow those AFP-positive relatives without overt tumors for the subsequent development of cancer.

Graft-Versus-Host Disease and Lymphomas

During chronic murine graft-versus-host (GVH) disease, as shown by Keller et al, approximately one third of the splenic lymphocytes demonstrated surface AFP by immunofluorescence (19). Splenic lymphocytes also have a suppressed PHA response, while lymph-node cells from the same GVH animal have no surface AFP and a normal mitogen response. Synthesis of AFP has been demonstrated in GVH spleens as well as in the in vitro allogeneic reaction (MLR), as measured by ^{14}C-leucine incorporation. GVH animals have normal levels of AFP by radioimmunoassay, demonstrating the phenomena of local (splenic) immune suppression associated with AFP-bearing cells and AFP synthesis in the absence of serum elevations. Local production of AFP is also shown by the finding of splenic AFP and suppression with only a small amount of AFP synthesis and no immunosuppression in lymph nodes. We believe that these observations could have importance in certain malignancies where regional immune suppression at the site of the tumor, which may not be reflected in studies of systemic immunity, may be operative. This is also the first demonstration of the synthesis of AFP in lymphoid tissues and has implications in theories relating oncogenesis and immunity. Since lymphoma cells could not be demonstrated in the spleens of the chronic GVH animals and AFP production was seen in in vitro allogeneic reactions (MLR), we postulate that embryonic antigens, including AFP, appear on the surface of cells undergoing transformation in response to allogeneic stimuli. AFP produced locally from blast cells may be of importance in the development of malignancies, either by suppressing immune surveillance against tumor cells or by suppressing immunity against oncogenic viruses.

Since GVH animals often manifest C-type RNA viruses and develop lymphomas, it was of interest to examine AFP in animals and humans with malignant lymphomas. Three murine B-cell lymphomas (L1210, P1798, C3HD) were shown to contain surface AFP-positive cells in their spleens, AFP synthesis by ^{14}C-leucine incorporation, and suppression of mitogen response. Serum levels of AFP were normal. Ascites cells from the L1210 and P1798 tumors, consisting of 90% tumor cells, did not synthesize or bind AFP in vitro. This suggests that the cells producing AFP and those with surface receptors in the spleens of lymphomatous animals were not the malignant lymphocytes. Whether the same or different cells both produce and bind AFP and the nature of the cell type(s) is unknown.

In a study of 10 patients with biopsy-proven malignant lymphomas, Keller et al (55) found that eight manifested AFP bound to mononuclear cells in involved lymphoid organs. The presence of AFP-bearing cells was associated with a diminished response to PHA and suppression was reversed by culturing in media alone for seven days. The culture supernates suppressed the mitogen response of normal mononuclear cells and the suppressive factor was removed by AFP affinity chromatography. Spleens and lymph nodes from AFP-positive patients synthesized AFP in vitro, while AFP-negative and control lymphoid tissues did not. These studies therefore suggest, but do not directly prove, that AFP is involved in the immunosuppression seen in lymphomatous disease. However, they shed no light on the role of suppression per se in the disease.

AFP in Diagnosis of Fetal Abnormalities

AFP may have diagnostic significance in certain fetal disorders. Elevations of AFP in maternal serum are seen in Rh-immunized and diabetic pregnancies and are associated with intrauterine fetal death. Maternal elevations are probably associated with increased feto-maternal transfusion of blood and in cases of intrauterine death with resorption of the fetus. Raised levels of amniotic-fluid AFP are seen in a variety of neural-tube defects and are important (in conjunction with sonar examination) in the antenatal diagnosis of anencephaly and spina bifida. Other defects associated with elevated amniotic-fluid AFP include hydrocephalus, meningocele, myelocele, Fallot's tetralogy, congenital nephrosis, esophageal atresia, and severe Rh disease (56).

CONCLUSION

AFP is one of several oncofetal proteins synthesized in large amounts by the fetus. Although synthesis drops markedly shortly after birth, small amounts of AFP continue to be produced in the adult. The function of AFP is unknown, but recent studies suggest the possibility that it may have immunoregulatory properties and/or may influence cell proliferation and growth. The high affinity of AFP for estrogen could have important biological functions, although the significance of this binding has not as yet been clearly defined.

Elevated levels of AFP are seen in a variety of clinical situations, including pregnancy; hepatic disorders, especially chronic hepatitis; and various malignancies, particularly hepatomas, teratomas, and those of primitive gut origin. It is also produced in murine GVH reactions and in lymphomas, both in mice and humans.

In human hepatic diseases, as well as the malignancies found in cancer-prone families, human and murine lymphomas, and murine GVH reactions, the presence of AFP-positive cells and immune suppression are highly correlated, but the role of these in the pathogenesis of the diseases is as yet unclear. It appears, however, that AFP may be produced locally in lymphoid tissues involved in GVH and lymphomatous disease without elevations in serum AFP levels. It is speculated that the local production of AFP in these situations may result from blastogenesis of lymphoid cells directed against foreign tumor or viral antigens. AFP could promote the development of tumors either by suppressing immune surveillance and/or immunity to oncogenic viruses, although this is speculative. Finally, AFP elevations in maternal serum and amniotic fluid are valuable diagnostically in the detection of fetal abnormalities, particularly neural-tube defects.

Literature Cited

1. Ruoslahti, E., Pihko, H., Seppala, M. 1974. Alpha-fetoprotein: Immunochemical purification and chemical properties. Expression in normal state and in malignant and nonmalignant liver diseases. *Transplant. Rev.* 20: 38–60
2. Gitlin, D. 1975. Normal biology of α-fetoprotein. *Ann. NY Acad. Sci.* 259: 7–16
3. Nishi, S., Hirai, H. 1973. Radioimmunoassay of α-fetoprotein in hepatoma, other liver diseases, and pregnancy. *Gann Monogr. Cancer Res.* 14: 79–87
4. McIntire, K. R., Waldmann, T. A., Moertel, C. G., Go, V.L.W. 1975. Serum alpha-fetoprotein in patients with neoplasms of the gastro-intestinal tract. *Cancer Res.* 35:991–96
5. Nishi, S., Watabe, H., Hirai, H. 1975. Immunological and chemical correlation between α-fetoproteins from human and several mammalian species. *Ann. NY Acad. Sci.* 259:109–18
6. Watanabe, A., Taketa, K., Kosaka, K. 1975. Microheterogeneity of rat α-fetoprotein. *Ann. NY Acad. Sci.* 259:95–108
7. Alpert, E., Perencevich, R. C. 1975. Human α-fetoprotein immunochemical analysis of isoproteins. *Ann. NY Acad. Sci.* 259:131–35
8. Ruoslahti, E., Terry, W. D. 1976. α-fetoprotein and serum albumin show sequence homology. *Nature* 260:804–5
9. Ruoslahti, E., Pihko, J. 1975. Effect of chemical modification on the immunogenicity of homologous α-fetoprotein. *Ann. NY Acad. Sci.* 259:85–94
10. Tsukada, Y., Hirai, H. 1975. α-fetoprotein and albumin synthesis during the cell cycle. *Ann. NY Acad. Sci.* 259:37–44
11. Sell, S., Morris, H. P. 1974. Relationship of rat α-1-fetoprotein to growth rate and chromosome composition of Morris hepatomas. *Cancer Res.* 34: 1413–17
12. Sell, S., Skelly, H., Leffert, H. L., Müller-Eberhard, U., Kida, S. 1975. Relationship of the biosynthesis of α-fetoprotein, albumin, hemopexin, and haptoglobin to the growth state of fetal rat hepatocyte cultures. *Ann. NY Acad. Sci.* 259:45–58
13. Becker, F., Sell, S. S. 1974. Early evaluation of α1-fetoprotein in N-2-fluorenylacetamid hepatocarcinogenesis. *Cancer Res.* 34:2489–94
14. Belanger, L., Hamel, D., Lachance, L., Dufour, D., Tremblay, M., Gagnon, P. M. 1875. Hormonal regulation of alphafoetoprotein. *Nature* 256:657–59
15. Leffert, H. L., Sell, S. 1974. Alpha$_1$-fetoprotein biosynthesis during the growth cycle of differentiated fetal rat hepatocytes in primary monolayer culture. *J. Cell Biol.* 61:823–29
16. Leffert, H. L. 1976. In *Workshop on Rat Liver Pathology, Woods Hole, Mass.*, ed. P. Newberne, W. Butler. Amsterdam: Elsevier. In press
17. Nunez, E. A., Sauv, L., Engelmann, F., Benassayag, C., Crepy, O., Jayle, M. F. 1971. Biochimie-origine embryonnaire de la proteine serique fixant l'aestrone et l'aestradiol chez la ratte impubere. *C. R. Acad. Sci.* 273D:242–45
18. Uriel, J., de Nechaud, B., Dupiers, M. 1972. Estrogen-binding properties of rat, mouse and man fetospecific serum proteins. Demonstration by immunoautoradiographic methods. *Biochem. Biophys. Res. Commun.* 46:1175–80
19. Keller, R. H., Tomasi, T. B. Jr. 1976. Alpha-fetoprotein synthesis by murine lymphoid cells in allogeneic reactions. *J. Exp. Med.* 143:1140–53
20. Uriel, J., Bouillon, D., Aussel, C., Dupiers, M. 1976. Alpha-fetoprotein: The major high-affinity estrogen binder in rat uterine cytosols. *Proc. Natl. Acad. Sci.* 73:1452–56
21. O'Malley, B. W., Means, A. R. 1974. Female steroid hormones and target cell nuclei. *Science* 183:610–20
22. Gitlin, D., Boesman, M. 1966. Serum α-fetoprotein, albumin, and γG-globulin in the human conceptus. *J. Clin. Invest.* 45:1826–38
23. Masseyeff, R., Gilli, J., Krebs, B., Calluaud, A., Bonet, C. 1975. Evolution of α-fetoprotein serum levels throughout life in humans and rats, and during pregnancy in the rat. *Ann. NY Acad. Sci.* 259:17–28
24. Gitlin, D., Kitzes, J., Boesman, M. 1967. Cellular distribution of serum α-foetoprotein in organs of the foetal rat. *Nature* 215:534
25. Murgita, R. A., Tomasi, T. B. Jr. 1975. Suppression of the immune response by α-fetoprotein. I. The effect of mouse α-fetoprotein on the primary and secondary antibody response. *J. Exp. Med.* 141:269–86
26. Murgita, R. A., Tomasi, T. B. Jr. 1975. Suppression of the immune response by

α-fetoprotein. II. The effect of mouse α-fetoprotein on mixed lymphocyte reactivity and mitogen-induced lymphocyte transformation. *J. Exp. Med.* 141:440–52

27. Ogra, S. S., Murgita, R. A., Tomasi, T. B. Jr. 1974. Immunosuppressive activity of mouse amniotic fluid. *Immunol. Commun.* 3(5):497–508
28. Caldwell, J. L., Severson, C. D., Thompson, J. S. 1973. Human alpha-fetoprotein: Embryonic alpha-1-globulin with in vitro immunosuppressive activity. *Fed. Proc.* 32:979 (Abstr.)
29. Parmely, M. J., Hsu, H. F. 1973. Rat alpha-fetoprotein: Inhibitory activity on lymphocyte cultures. *Fed. Proc.* 32:979 (Abstr.)
30. Belanger, L., Waithe, W. I., Daguillard, F., Larochelle, J., Dufour, D. 1974. The effect of AFP on the in vitro lymphocyte response. In *L'Alpha-foetoprotein,* ed. R. Masseyeff, pp. 423–30. Paris: Inserm
31. Yachnin, S. 1976. Demonstration of the inhibitory effect of human alpha-fetoprotein on in vitro human lymphocyte transformation. *Proc. Natl. Acad. Sci.* In press
32. Yachnin, S., Lester, E. Inhibition of human lymphocyte transformation by human alpha-fetoprotein (HAFP): Comparison of fetal and hepatoma HAFP and kinetic studies of in vitro immunosuppression. *Clin. Exp. Immunol.* In press
33. Sheppard, H. W. Jr., Sell, S. 1976. Effect of AFP on immune phenomona in rats and mice. Presented at Conf. Onco-Dev. Gene Expression, San Diego, May 1976. In press
34. Belanger, L., Larochelle, J., Daguillard, F., Belanger, M. Is AFP an immune regulator? See Ref. 33
35. Goeken, N. E., Thompson, J. S. Immunosuppressive properties of human alpha-fetoprotein and human cord plasma albumin. See Ref. 33
36. Tyan, M. L. 1976. Immunosuppressive properties of mouse amniotic fluid. *Proc. Soc. Exp. Biol. Exp. Med.* 151:343–47
37. Yachnin, S. 1975. Fetuin, an inhibitor of lymphocyte transformation. *J. Exp. Med.* 141:242–56
38. Calvanico, N. J., Dattwyler, R. J., Allen, J., Tomasi, T. B. Jr. 1976. The effect of bovine amniotic fluid proteins on the generation of murine cytotoxic T cells. *Fed. Proc.* 32:250 (Abstr.)
39. Dattwyler, R. J., Murgita, R. A., Tomasi, T. B. Jr. 1975. Binding of α-fetoprotein to murine T cells. *Nature* 256:656–57
40. Keller, R. H., Tomasi, T. B. Jr. 1975. Alpha-fetoprotein, rosette formation and tumor cell growth. *N. Engl. J. Med.* 293:1264–65
41. Baumgarten, A. 1976. E rosette formation in the presence of α-fetoprotein and ferritin. *Clin. Immunol. Immunopathol.* 6:42–46
42. Uriel, J., Aussel, C., Bouillon, D., Loisillier, F., de Nechaud, B. 1975. Liver differentiation and the estrogen-binding properties of α-fetoprotein. *Ann. NY Acad. Sci.* 259:119–30
43. Keller, R. H., Calvanico, N. J., Tomasi, T. B. Jr. Immunosuppressive properties of AFP: Role of estrogens. See Ref. 33
44. Sterzle, J., Silverstein, A. M. 1967. Developmental aspects of immunity. *Adv. Immunol.,* ed. F. J. Dixon, J. H. Humphrey, 6:337–459. New York: Academic
45. Warner, N. L., Herzenberg, L. A. 1970. Tolerance and immunity to maternally derived incompatible IgG$_{2a}$-globulin in mice. *J. Exp. Med.* 132:440–47
46. Boehmer, H. V., Byrd, W. J. 1972. Responsiveness of thymus cells to syngeneic and allogeneic lymphoid cells. *Nature New Biol.* 235:50–52
47. Hellström, K. E., Hellström, I. 1972. The role of serum factors (blocking antibodies) as mediators of immunological nonreactivity to cellular antigens. In *Ontogeny of Acquired Immunity,* ed. R. Porter, J. Knight, pp. 133–47. *Ciba Found. Symp.* Amsterdam: Elsevier
48. Tomasi, T. B. Jr., Murgita, R. A., Thompson, R. L., Calvanico, N. J., Dattwyler, R. J. 1976. Alpha-fetoprotein and the immune response during murine pregnancy and neonatal development. Presented at Natl. Inst. Child Health Hum. Dev. Conf. Dev. Host Def., Eldridge, Maryland, May 1976. In press
49. Strelkauskas, A. J., Wilson, B. S., Dray, S., Dodson, M. 1975. Inversion of levels of human T and B cells in early pregnancy. *Nature* 258:331–32
50. Slade, B. 1973. Antibodies to α-foetprotein cause foetal mortality in rabbits. *Nature* 256:493–94
51. Dudley, F. J., Fox, R. A., Sherlock, S. 1972. Cellular immunity and hepatitis-associated Australian antigen liver disease. *Lancet* 1:723–26

52. Keller, R. H., Summerskill, W. H. J., Geubel, A. P., Shorter, R. G., Tomasi, T. B. Jr. 1976. Alpha-fetoprotein and suppression of cell-mediated immunity in chronic active liver disease (CALD). *Gastroenterology* 70:A-43–901 (Abstr.)
53. Lynch, H. T. 1976. *Cancer Genetics.* Springfield, Ill: Thomas. 639 pp.
54. Keller, R. H., Guirgis, H., Lynch, H., Tomasi, T. B. Jr. Alpha-fetoprotein (AFP) in cancer-prone families. *Clin. Res.* 24:331 (Abstr.)
55. Keller, R. H., Dattwyler, R. J., Harrison, E., Tomasi, T. B. Jr. 1975. Alpha-fetoprotein in human lymphoproliferative diseases. *Blood* 46(6):1035
56. Seppala, M. 1975. Fetal pathophysiology of human α-fetoprotein. *Ann. NY Acad. Sci.* 259:59–73

Ann. Rev. Med. 1977. 28:467–73
Copyright © 1977 by Annual Reviews Inc. All rights reserved

PATHOPHYSIOLOGY OF PSORIASIS[1]

❖7255

John J. Voorhees, M.D.
Department of Dermatology, University of Michigan Medical School,
Ann Arbor, Michigan 48109

INTRODUCTION

The best available data indicate that psoriasis is a multifactorial genetic disease (1) with a three- to fourfold increased frequency of the cell-surface gene products HL-A13 and HL-A17 (2). Individuals with psoriasis genes develop the disease either spontaneously or in an area of damage to the skin. This disease is characterized by uncontrolled, nonmalignant growth of the epithelium (epidermis) of skin. Diseased epidermis is characterized by markedly increased proliferation (3), reduced terminal differentiation (4), and elevated glycogen content (4). Without uncontrolled epidermal growth, the disease would probably not be clinically significant. Therefore the majority of psoriasis research has focused on the molecular control of epidermal growth and differentiation. This focus seems especially appropriate, since the normal-appearing skin of the psoriatic patient is certainly not normal. Although not sufficient to reach the clinical horizon, growth is only partially regulated in clinically normal areas. The evidence for this is a twofold increase in proliferating cells in comparison to normal volunteers (5).

Psoriasis is a complex disease composed of several compartments that have been reviewed previously (6). This paper reviews facts on the control of epithelial growth in psoriasis available in September, 1976. The aim is to discuss pertinent molecular pathophysiology. Aside from the obvious genetic component, the ultimate cause of unrestrained epidermal growth is unknown and is not considered. Rather, the goal is to identify those factors, among the myriad of growth-related events, that might be of signal importance to growth control, or that, if modulated pharmacologically, might normalize growth. Space limitations require that many references in this paper be to exhaustive reviews that we previously published. We apologize to authors whose pioneering work has not been cited here, but in all cases, the primary publication can be located in the referenced review.

---

[1] Supported in part by Program Project (2 PO1 AM 15740–05) "Mechanisms and Therapy of Psoriasis" from the National Institute of Arthritis, Metabolism, and Digestive Diseases, National Institute of Health, Bethesda, Maryland.

CELL SURFACE IN PSORIASIS

The cell surface is thought to participate in the regulation of growth and differentiation of most animal cells. Ultrastructural analysis reveals almost complete absence of the lesional epidermal cell surface (7). This may be a general feature of proliferating epithelium, nonspecific for psoriasis. Nonspecificity should not diminish the importance of this observation. It is possible that normal epidermal growth regulation cannot occur without a normal cell surface (7). The *HL-A* gene products are cell-surface constituents and, as mentioned above, are abnormal in psoriasis (2). Whether these HL-A abnormalities are genetic markers or participants in the pathophysiology of the disease is unknown.

The *H-2* gene complex of the mouse is analogous to the *HL-A* gene complex in man. The *H-2* genotype is one regulator of the levels of liver cyclic AMP (cAMP) in mice (8). Presumably the *H-2* genotype controls the cell-surface membrane phenotype, and by analogy the *HL-A* genotype in man may do likewise. The increased frequency of HL-A13 and HL-A17 cell-surface components in psoriasis could prevent normal cell-surface regulation of the cAMP system. The abnormal HL-A psoriatic phenotype could alter cell adhesion–mediated growth control. Alternatively, this phenotype may interfere with hormonal regulation of epidermal cAMP levels. If beta catecholamines regulate epidermal cAMP content physiologically, the HL-A13 and HL-A17 psoriatic cell-surface phenotype might account for the observed practolol (a beta-blocking agent)-induced, psoriasis-like rashes (9). As we show below, we believe that reduced function of the epidermal cAMP system is central to the uncontrolled growth of the psoriatic epidermis (4).

Cellular cAMP content is regulated by unknown cell-surface constituents that may not be HL-A antigens, but may be hormone receptors or cell-cell recognition sites (4). Thus the reported cell-surface defects in psoriasis (7) could reduce epidermal cAMP function without participation of the abnormal HL-A system. It is known that reduced cAMP function can alter cell-surface constituents (10). Thus the cell surface regulates cAMP, which may in turn regulate the cell surface. The cell surface and cAMP are critical regulators of cell physiology, which, in epidermis, is the maintenance of normal growth (4). Therefore, this discussion of psoriasis focuses on effectors and events associated with the cell surface and cyclic nucleotides.

CYCLIC AMP AND CYCLIC GMP IN PSORIASIS

The first working model of cAMP in psoriasis was put forward by Voorhees & Duell in 1971 (11). It was known that three characteristic features of the lesional epidermis were increased proliferation (3), decreased terminal differentiation (4), and glycogen accumulation (4). In other tissues and experimental systems, cAMP was capable of normalizing these three features (4). In the model, reduced cAMP function could either trigger the onset of the lesion, permit a lesion to be maintained, or both (11). Since the absence of these three features describes a normal epidermis, a prominent feature of the model is that cAMP is necessary for normal epidermal function, a major feature of which is regulated growth. This model was modified in 1973 to accommodate the observation that elevated cellular levels of cyclic GMP (cGMP) are associated with stimulated lymphocyte proliferation. The present working model suggests that reduced cAMP function, increased cGMP function, or altered function associated with a decrease in the cAMP/cGMP ratio are central to lesional pathophysiology. Since 1971 these models have been extensively studied in several laboratories, including our own. The results and our interpretation of these investigations are summarized here.

We reported a significantly lower level of cAMP in involved epidermis (IE) versus uninvolved epidermus (UE), in 50 patients using two different analytical methods to estimate cAMP content (4). In 1974 Härkönen et al (12) found cAMP levels to be twice higher in IE than epidermis from normal volunteers. It is difficult to compare this study with our earlier work, since the cAMP assay was different. However, more important is the fact that this study measured cAMP levels in the outer half of lesions, whereas we routinely used the entire lesional epidermis for all comparisons with uninvolved or normal epidermis. Consequently, these data are very difficult to interpret. In 1975 Yoshikawa et al (13), using micromethodology, found the cAMP content of IE versus UE to be essentially the same. More recently we have used a sensitive and specific radioimmunoassay to measure cAMP and cGMP levels simultaneously in IE versus UE from another sixteen patients. In contrast to our previous work, we found the mean cAMP levels in IE versus UE to be essentially the same (C. Marcelo, E. A. Duell, T. F. Anderson, M. A. Stawiski, J. J. Voorhees, manuscript in preparation), in agreement with Yoshikawa et al (13). In our previous work, which demonstrated significantly lower mean cAMP levels in IE versus UE, a wide scatter of cAMP values from patient to patient was a constant feature. In summary, the levels of cAMP in IE versus UE do not appear higher in IE and they may be lower (4). The data of Yoshikawa et al (13) and our own recent data discussed above indicate that the levels of cAMP in IE versus UE are probably the same. We believe that the scatter of the data and the relatively small observable changes suggest that further study of cAMP levels per se will be unrewarding.

Since a change in cAMP levels may be more important than the steady-state level, the stimulability of cAMP formation has been studied. The IE appears to accumulate less cAMP in response to PGE and beta catecholamine than does UE (14, 15). Our work indicates that cAMP accumulates to a similar level after exposure to beta catecholamine in IE versus UE (4). This suggests that the beta receptor and adenylate cyclase have normal activity. However, Mui et al (16) have data that suggest reduced adenylate cyclase activity in IE versus UE. For technical reasons, neither our work (4) nor that of Mui et al (16) is sufficiently rigorous to judge the activity of adenylate cyclase in IE. In preliminary results, which require confirmation, the low-K_m form of cAMP phosphodiesterase is more active in IE versus UE in six patients (17). Cyclic AMP–dependent protein kinase has not yet been examined in a manner sufficient to permit any conclusions. A rise in lesional cAMP reduces lesional glycogen content (18). The only data on GMP metabolism has come from our laboratory. We initially reported a significant mean increase in cGMP levels in IE versus UE. Subsequently, by radioimmunoassay we again found a significant increase in the cGMP levels of IE versus UE (C. Marcelo, E. A. Duell, T. F.

Anderson, M. A. Stawiski, J. J. Voorhees, manuscript in preparation). Since cAMP and cGMP were assayed in the same samples, the cAMP/cGMP ratio is decreased due to the increase in the cellular content of cGMP.

From the above, it is clear that much more must be done to clarify the role of cyclic nucleotides in psoriasis. There is general agreement that the cAMP system is deranged in psoriasis and that cAMP function is important in psoriasis. However, from the above discussion there is little agreement on the site of the defect. The cAMP/cGMP ratio is reduced and this may be important. Its true importance will not be understood until the function of cGMP in epidermis is known.

CYCLIC AMP IN EPIDERMAL FUNCTION

Cyclic AMP is thought to be important in the pathophysiology of psoriasis for several reasons. Lithium therapy is capable of exacerbating psoriasis, and lithium is known to inhibit cAMP formation in several tissues, as well as in psoriatic epithelium in vitro (19). Practolol can cause psoriasis-like rashes. The ability of practolol to interfere with cAMP formation may permit psoriasis to develop in a genetically predisposed individual (9). Conversely, papaverine and RO-20-1724 are known to promote the accumulation of cAMP in epidermis in vitro and to improve psoriasis when used clinically (19, 20). Cyclic AMP inhibits the proliferation of epidermal cells in vitro (19) and induces differentiation of epidermis in vitro (21).

In wound healing, the epidermis exhibits increased proliferation, decreased differentiation, and glycogen accumulation in a manner similar to psoriasis. This is prevented or reversed by elevating cAMP in the wounded epithelium (22). The skin tumor–promotion phase of two-stage carcinogenesis is characterized by epithelial proliferation and reduced differentiation. This process has been studied extensively by Marks (23), who believes that the relative tissue levels of cAMP and cGMP control the equilibrium between proliferation and differentiation. In IE, polyamine levels are significantly higher than in UE (24). Since cAMP is capable of stimulating polyamine synthesis in certain tissues (25), it is possible that elevated polyamine formation can be inhibited by cAMP in other tissues such as epidermis. Thus, although cyclic nucleotide metabolism in psoriasis is poorly understood, it seems possible that future studies in this area will improve our understanding of this disease.

REGULATION OF EPIDERMAL CYCLIC NUCLEOTIDE METABOLISM IN PSORIASIS

In spite of extensive study, we have been unable to identify factors that consistently exhibit the ability to regulate epidermal cGMP levels (C. Marcelo, unpublished data). Prostaglandins E_1 and E_2, beta catecholamines, and adenosine (26), as well as histamine (27), can alter epidermal cAMP content. Psoriasis occurs in discrete patches and wound healing is obviously under local control. Prostaglandin is known to act over a very short distance and thus may be the best candidate for the physiological regulation of epidermal cAMP content.

PROSTAGLANDINS AND PSORIASIS

Prostaglandins are formed from free arachidonic acid, which in normal tissue exists in very minute amounts. In psoriasis the content of free arachidonic acid in IE is 26 times greater than in UE. Prostaglandins E_2 and $F_{2\alpha}$ are increased by only 40% and 86% in IE and UE, respectively. Another metabolite of free arachidonic acid, 12L-hydroxy-5, 8, 10, 14-eicosatetraenoic acid (HETE), is increased 82-fold in IE versus UE. Clearly the majority of free arachidonic acid is being converted to HETE rather than to the classic prostaglandins or thromboxanes (28). One possible explanation is the inhibitor(s) of prostaglandin biosynthesis, which Penneys et al have detected in homogenates of involved psoriatic epithelium (29). The pathophysiological relevance of the abovementioned abnormalities in arachidonic acid metabolism are unknown. However, it was recently demonstrated that glucocorticoids markedly inhibit the formation of arachidonic acid (30). According to current understanding of arachidonic acid metabolism, glucocorticoid would thus depress not only the formation of arachidonic acid, but also that of HETE, thromboxane, PGE_2 and $PGF_{2\alpha}$. Topical glucocorticoids are the most potent antipsoriatic agents known. It is possible that inhibiting the formation of arachidonic acid and its metabolites is one reason for the efficacy of topical glucocorticoids in psoriasis.

POLYAMINES AND PSORIASIS

As previously mentioned, polyamine biosynthesis can be regulated by cAMP (25). Stimulation of the rate of biosynthesis of the polyamines spermidine and spermine appears to be the most constant correlate of increased cell proliferation (31). We have recently reported that the level of polyamines is significantly higher in IE versus UE and that the polyamine content of UE is significantly higher than in epidermis from control volunteers. This is probably a molecular correlate of the twofold increase in the labeling index of UE versus epidermis from control volunteers (5). Putrescine is the precursor of spermidine (31). Putrescine is formed from ornithine by the action of ornithine decarboxylase (31). Of interest is the fact that several years ago Rothberg & Van Scott showed that arginase activity was markedly elevated in psoriatic lesions. Arginase is responsible for the formation of ornithine from arginine (32).

Although elevated polyamine levels have been associated with rapid proliferation for several years (31), it has been only recently that a potential cause-and-effect relation between polyamines and proliferation has been suggested. Mamont et al recently showed that blocking the synthesis of polyamines by an inhibitor of ornithine decarboxylase could reduce cell proliferation (33). This inhibition could be prevented by combining polyamines with the blocking agent, thus bypassing the ornithine decarboxylase blockade. Thus it appears that regulating polyamine synthesis in psoriasis by cAMP or other agents might be productive. This seems especially desirable in view of a recent report that shows that polyamine concentrations comparable to those we detected in psoriatic lesions can inhibit the ability of cAMP to stimulate epidermal protein kinase (34). The high polyamine content of a psoriatic lesion could reduce the function of the cAMP system in vivo, thus accounting for all changes predicted by the cyclic nucleotide model of psoriasis without any detectable defect in the cAMP system as assayed in vitro.

SUMMARY

Psoriasis is a disease characterized by benign but unrestricted epithelial growth. The molecular pathophysiology of this unregulated growth has been discussed. A better understanding of the role of the cell surface, cyclic nucleotides, arachidonic acid–prostaglandin metabolism, and polyamines in psoriasis may clarify our understanding of this disease and produce clues pertinent to a general understanding of growth regulation.

Literature Cited

1. Watson, W., Cann, H. M., Farber, E. M., Nall, M. L. 1972. The genetics of psoriasis. *Arch. Dermatol.* 105:197–207
2. Svejgaard, A., Platz, P., Ryder, L. P., Nielsen, L. S., Thomsen, M. 1975. HL-A and disease associations—a survey. *Transplant. Rev.* 22:3–43
3. Weinstein, G. D., Frost, P. 1971. Methotrexate for psoriasis: A new therapeutic schedule. *Arch. Dermatol.* 103: 33–38
4. Voorhees, J. J., Duell, E. A., Stawiski, M., Harrell, E. R. 1974. Cyclic nucleotide metabolism in normal and proliferating epidermis. In *Advances in Cyclic Nucleotide Research,* ed. P. Greengard, G. A. Robison, 4:117–62. New York: Raven
5. Goodwin, P., Hamilton, S., Fry, L. 1973. A comparison between DNA synthesis and mitosis in uninvolved and involved psoriatic epidermis and normal epidermis. *Br. J. Dermatol.* 89:613–18
6. Farber, E. M., Cox, A. J., eds. 1971. *Psoriasis.* Stanford, Calif: Stanford Univ. Press
7. Mercer, E. H., Maibach, H. I. 1968. Intercellular adhesion and surface coats of epidermal cells in psoriasis. *J. Invest. Dermatol.* 51:215–21
8. Meruelo, D., Edidin, M. 1975. Association of mouse liver adenosine 3':5'-cyclic monophosphate (cyclic AMP) levels with histocompatibility-2 genotype. *Proc. Natl. Acad. Sci. USA* 72: 2644–48
9. Wadskov, S., Søndergaard, J., Svejgaard, A., Mikkelsen, H. I., Jensen, H. A. E. 1976. HL-A histocompatibility antigens and cutaneous reactions to practolol. *Acta Derm. Venereol.* 56: 161–62
10. Baig, M. M., Roberts, R. M. 1973. Comparative studies on the carbohydrate-containing membrane components of normal and adenosine 3':5'-cyclic monophosphate-treated Chinese-hamster ovary cells. *Biochem. J.* 134:329–39
11. Voorhees, J. J., Duell, E. A. 1971. Psoriasis as a possible defect of the adenyl cyclase-cyclic AMP cascade: A defective chalone mechanism? *Arch. Dermatol.* 104:352–58
12. Härkönen, M., Hopsu-Havu, V. K., Raij, K. 1974. Cyclic adenosine monophosphate, adenyl cyclase and cyclic nucleotide phosphodiesterase in psoriatic epidermis. *Acta Derm. Venereol.* 54:13–18
13. Yoshikawa, K., Adachi, K., Halprin, K. M., Levine, V. 1975. Is the cyclic AMP in psoriatic epidermis low? *Br. J. Dermatol.* 93:253–58
14. Adachi, K., Yoshikawa, K., Halprin, K. M., Levine, V. 1975. Prostaglandins and cyclic AMP in epidermis. *Br. J. Dermatol.* 92:381–88
15. Aso, K., Orenberg, E. K., Farber, E. M. 1975. Reduced epidermal cyclic AMP accumulation following prostaglandin stimulation: Its possible role in the pathophysiology of psoriasis. *J. Invest. Dermatol.* 65:375–78
16. Mui, M. M., Hsia, S. L., Halprin, K. M. 1975. Further studies on adenyl cyclase in psoriasis. *Br. J. Dermatol.* 92:255–62
17. Voorhees, J. J., Colburn, N. H., Stawiski, M., Duell, E. A., Haddox, M., Goldberg, N. D. 1974. Imbalanced cyclic AMP and cyclic GMP levels in the rapidly dividing, incompletely differentiated epidermis of psoriasis. In *Control of Proliferation in Animal Cells,* ed. R.

Baserga, B. Clarkson, pp. 635–48. New York: Cold Spring Harbor Lab.

18. Halprin, K., Adachi, K., Yoshikawa, K., Levine, V., Mui, M. M., Hsia, S. L. 1975. Cyclic AMP and psoriasis. *J. Invest. Dermatol.* 65:170–78
19. Voorhees, J. J., Marcelo, C. L., Duell, E. A. 1975. Cyclic AMP, cyclic GMP, and glucocorticoids as potential metabolic regulators of epidermal proliferation and differentiation. *J. Invest. Dermatol.* 65:179–90
20. Stawiski, M., Rusin, L., Schork, M. A., Burns, T., Duell, E., Voorhees, J. 1976. RO 20-1724 elevates epidermal cyclic AMP levels in vitro and improves psoriasis in vivo. *Clin. Res.* 24:267 (Abstr.)
21. Delescluse, C., Fukuyama, K., Epstein, W. L. 1976. Dibutyryl cyclic AMP-induced differentiation of epidermal cells in tissue culture. *J. Invest. Dermatol.* 66:8–13
22. Williams, J. P. G. 1972. Interrelation of epithelial glycogen, cell proliferation and cellular migration with cyclic adenosine monophosphate in epithelial wound healing. *Cell Differ.* 1:317–23
23. Marks, F. 1976. Epidermal growth control mechanisms, hyperplasia, and tumor promotion in the skin. *Cancer Res.* 36:2636–43
24. Russell, D. H., Stawiski, M. A., Duell, E. A. Voorhees, J. J. 1976. Elevated polyamine metabolism in patients with psoriasis. *Pharmacologist* 18(2):157 (Abstr.)
25. Byus, C. V., Russell, D. H. 1974. Effects of methyl xanthine derivatives on cyclic AMP levels and ornithine decarboxylase activity of rat tissue. *Life Sci.* 15:1991–97
26. Voorhees, J. J., Duell, E. A. 1975. Imbalanced cyclic AMP-cyclic GMP levels in psoriasis. In *Advances in Cyclic Nucleotide Research,* ed. G. I. Drummond, P. Greengard, G. A. Robison, 5:735–58. New York: Raven
27. Iizuka, H., Adachi, K., Halprin, K. M., Levine, V. 1976. Histamine (H_2) receptor-adenylate cyclase system in pig skin (epidermis). *Biochim. Biophys. Acta* 437:150–57
28. Hammarström, S., Hamberg, M., Samuelsson, B., Duell, E. A., Stawiski, M., Voorhees, J. J. 1975. Increased concentrations of nonesterified arachidonic acid, 12L-hydroxy-5,8,10,14-eicosatetraenoic acid, prostaglandin E_2, and prostaglandin $F_{2\alpha}$ in epidermis of psoriasis. *Proc. Natl. Acad. Sci. USA* 72:5130–34
29. Penneys, N. S., Ziboh, V., Lord, J., Simon, P. 1975. Inhibitor(s) of prostaglandin synthesis in psoriatic plaque. *Nature* 254:351–52
30. Hong, S. L., Levine, L. 1976. Inhibition of arachidonic acid release from cells as the biochemical action of anti-inflammatory corticosteroids. *Proc. Natl. Acad. Sci. USA* 73:1730–34
31. Fillingame, R. H., Jorstad, C. M., Morris, D. R. 1975. Increased cellular levels of spermidine or spermine are required for optimal DNA synthesis in lymphocytes activated by concanavalin A. *Proc. Natl. Acad. Sci. USA* 72:4042–45
32. Rothberg, S., Van Scott, E. J. 1958. Evaluation of arginase activity in normal epidermal tissue and pathological stratum corneum. *J. Invest. Dermatol.* 31:263–68
33. Mamont, P. S., Böhlen, P., McCann, P. P., Bey, P., Schuber, F., Tardif, C. 1976. α-Methyl ornithine, a potent competitive inhibitor of ornithine decarboxylase, blocks proliferation of rat hepatoma cells in culture. *Proc. Natl. Acad. Sci. USA* 73:1626–30
34. Murray, A. W., Froscio, M., Rogers, A. 1976. Effect of polyamines on cyclic AMP-dependent and independent protein kinases from mouse epidermis. *Biochem. Biophys. Res. Commun.* 71:1175–81

Ann. Rev. Med. 1977. 28:475–88
Copyright © 1977 by Annual Reviews Inc. All rights reserved

SURGICAL ADJUVANT CHEMOTHERAPY

❖7256

Marjorie Perloff, M.D. and James F. Holland, M.D.
Department of Neoplastic Diseases, Mount Sinai School of Medicine
of the City University of New York, New York, NY 10029

INTRODUCTION

Over the past 15 years, information has accumulated from experimental animal systems showing that surgical adjuvant chemotherapy produces a longer disease-free interval, an increased survival time, and a higher percentage of cures than either surgery or chemotherapy alone. An inverse relationship exists between the number of solid tumor cells and the chemotherapeutic response obtained (1). When all visible tumor is removed, as opposed to subtotal resection, the ability of chemotherapy to produce cure is greater (2). Micrometastases of 10^6 cells or less are clinically undetectable and even larger metastatic tumors may go unrecognized. Knowledge of the natural history of such micrometastases evolving into manifest disease makes them the target for surgical adjuvant therapy. Current knowledge of tumor-cell populations indicates that the growth fraction is inversely related to population size, that the generation time of tumor cells in small populations is often shorter than in large populations, and that first-order cell kinetics characterize drug kill of tumor cells (3). The evidence that these principles also apply to a variety of neoplastic diseases in man is discussed below.

WILMS' TUMOR

Reported survival rates for Wilms' tumor in the United States have increased substantially over the last two decades. As many as 80% of children presenting with localized disease and about 50% presenting with metastatic disease can be cured with currently available therapy (4). Improvements in survival have occurred in all stages of disease and appear to result from better therapy rather than from earlier detection (5). Surgical removal alone of localized primary tumors produced a cure rate as high as 30% (6). Postoperative radiotherapy to the tumor bed in patients with localized disease increased survival rates to 50% (7). The natural history of

Wilms' tumor prior to the use of chemotherapy implies that microscopic metastatic disease was present at the time of diagnosis. After surgery and radiotherapy in children apparently free of metastatic disease, actinomycin D used in courses every three months for a period of two years resulted in 89% disease-free survival at two years (8). Subsequently, vincristine was shown to be an effective chemotherapeutic agent (9). In patients with Groups II or III disease (Group II: tumor extends beyond the kidney, but is completely resected; Group III: residual nonhematogenous tumor confined to abdomen), the combination of these two agents is significantly superior to either agent alone in producing two-year disease-free survival (79% vs 50%) (10). For tumor confined exclusively to the kidney and completely resected (Stage I), surgery plus actinomycin D chemotherapy achieved the same results as surgery plus radiotherapy and chemotherapy (11). The second National Wilms' Tumor Study is presently evaluating whether 6 or 15 months of chemotherapy is superior for curing micrometastatic disease, and whether adriamycin in addition to actinomycin D and vincristine adds to the therapeutic regimen.

EWING'S SARCOMA

Ewing's sarcoma of bone has traditionally been a rapidly and almost uniformly fatal tumor (12). Surgical treatment alone has been of little value (13). Even if the primary tumor was successfully controlled by adequate irradiation, metastases to lung and bone frequently ensued, and 75% of patients were dead within two years of diagnosis (13). The combined administration of radiotherapy and systemic chemotherapy was first reported in 1962 (14). Adjuvant chemotherapy has been shown to be effective in decreasing the local recurrence rate and in delaying the median time to first metastasis (15). Rosen et al (16) report 12 children treated with radiation therapy and sequential adjuvant chemotherapy with actinomycin D, adriamycin, vincristine, and cyclophosphamide for two years. All remained without evidence of disease with follow-up periods ranging from 10 to 37 months. Pomeroy & Johnson reported 66 consecutive patients treated with local irradiation of the primary site combined with adjuvant regimens of progressively more intensive systemic chemotherapy. Most recently they have used alternating high-dose pulses of adriamycin and cyclophosphamide with vincristine. Forty-three patients without clinically detectable metastatic disease at the time of diagnosis had a 64% two-year and 52% five-year survival rate (17). Under the direction of the Intergroup Ewing's Sarcoma Study, a multi-institutional program has been undertaken. They have demonstrated the effectiveness of postradiation chemotherapy with cyclophosphamide, actinomycin D, and vincristine when compared to a control group without chemotherapy. To date, adriamycin has not been of significant clinical benefit when added to the combination chemotherapy following postoperative irradiation. However, additional bilateral pulmonary irradiation appears to increase the disease-free status of patients. (M. Nesbitt, personal communication). A current study tests whether adriamycin can achieve this effect without pulmonary irradiation, with its known compromise of pulmonary function years later.

CHILDHOOD EMBRYONAL RHABDOMYOSARCOMA

The survival of children with embryonal rhabdomyosarcoma has significantly improved in recent years. Neither extirpative surgery nor high-dose radiotherapy resulted in an overall survival of more than 25% (18). The addition of intensive maintenance chemotherapy to extirpative surgery and radiotherapy has resulted in a survival in the range of 60–80% (18, 19). Initially, single-agent chemotherapy was employed. The most beneficial agents included actinomycin D, vincristine, and cyclophosphamide (20, 14). Preservation of function and appearance in situations where these were formerly sacrificed has been made possible by advances in combination chemotherapy and changes in surgical thinking. Conservative trends that have evolved include the rarity of amputation for primary extremity tumors, possible irradiation for orbital tumors to obviate enucleation, avoidance of mutilating excision for head and neck tumors, and anterior exenteration for primary tumors of the bladder or prostate (21). Combination chemotherapy with actinomycin D, vincristine, cyclophosphamide, and adriamycin for 24 months, in addition to surgery and radiotherapy (4500–7000 rads), has resulted in an 82% disease-free interval ranging from 4 to 42+ months among children with Stages I–III disease (18).

OSTEOGENIC SARCOMA

The five-year survival rate in large series of patients treated for osteogenic sarcoma with ablation of the primary tumor either by surgery alone, or with radiotherapy directed to the primary site followed by surgery, is in the range of 20% (22, 23). X-ray evidence of pulmonary metastases occurs at a median of 8.5 months after potentially curative amputation and patients are usually dead within six months, implying that hematogenous pulmonary micrometastases are present in the majority of patients at the time of diagnosis (23, 24). In 1972 it was reported that both adriamycin and high-dose methotrexate followed by citrovorum rescue produced objective responses in children with metastatic osteogenic sarcoma (25, 26). Subsequent studies have employed these drugs alone or in various combinations in a surgical adjuvant fashion. When six monthly courses of adriamycin were given following radical surgical amputation, 71% of treated patients were reported to be free of pulmonary metastases as determined by life-table analysis at 18 months (27). More recent analysis of the study reveals a plateau at 75% survival, with observation times to four years among patients treated exactly according to the protocol (Cancer and Leukemia Group B, unpublished data). At the same time, it was reported that surgical adjuvant treatment with vincristine, high-dose methotrexate, and citrovorum rescue given for two years produced similar disease-free and survival intervals (28). Other studies have been reported employing adriamycin or high-dose methotrexate followed by citrovorum rescue in various combinations with vincristine, cyclophosphamide, and L-phenylalanine mustard. Among the 96 patients entered on these studies, the median disease-free interval ranges from 7 to 18 months (29–32).

Based on the above data in which the one- and two-year disease-free survival rates in patients with osteogenic sarcoma range from 50 to 80%, which is substantially different from surgery alone, attempts have been made to salvage weight-bearing limbs. Fifteen patients with biopsy-proven osteogenic sarcoma of the distal femur or proximal tibia were treated preoperatively with high-dose methotrexate and citrovorum rescue, as well as with adriamycin. This treatment was followed by en bloc excision of the primary tumor with prosthetic replacement, and postoperatively with high-dose methotrexate with citrovorum rescue, adriamycin, and high-dose cyclophosphamide given sequentially for one year. With follow-ups ranging from 2–15 months postoperatively, only one patient has relapsed either locally or distally (33). The demonstration that aggressive chemotherapy can cause destruction and regression of primary osteogenic sarcoma, and the proven efficacy of surgical adjuvant chemotherapy in this disease, provide a basis for the development of treatment regimens that should provide cure in a high percentage of patients and perhaps ultimately preserve involved limbs. One danger is the premature widespread adoption of such programs before the requisite drug regimens are clarified and appropriate surgical and medical teams assembled.

BREAST CANCER

Approximately 90,000 American women will develop breast carcinoma this year, and it is predicted that 47% of these will have axillary metastases at the time of diagnosis (34). By ten years, 24% of women with negative axillary nodes at the time of mastectomy have already relapsed. Among those with 1–3 positive nodes, the ten-year relapse rate is 65% and among those with four or more positive nodes it is 86% (35). Once metastatic disease develops, early death is virtually inevitable.

The initial surgical adjuvant trial in breast carcinoma, Stage I or II, was carried out by the National Surgical Adjuvant Breast Project. Patients were randomized to receive thiotepa or placebo injections at the time of operation and on the first two postoperative days. Among the 820 patients followed for more than five years and analyzed in this trial, thiotepa, as given, was of benefit only to premenopausal women with four or more positive lymph nodes at the time of mastectomy. The survival of these patients was 57% at five years vs 24% in the control group (36). Multiple surgical adjuvant chemotherapy trials were carried over the ensuing 15 years, but in none was the study design such that the results were conclusive.

Early studies demonstrated the efficacy of the five-drug Cooper regimen of vincristine (V), prednisone (P), cyclophosphamide (C), methotrexate (M), and 5-fluorouracil (F) in women with metastatic disease. Although this regimen was superior to the single-drug therapy previously employed, 75% of patients were dead within two years (37). Subsequently, a three-drug combination (CMF) was shown to be more effective than a single oral drug, L-phenylalanine mustard (L-PAM), in the treatment of women with metastatic breast carcinoma (38). Both of these treatments were then employed in controlled clinical trials, women with one or more positive axillary nodes at the time of mastectomy receiving either chemotherapy or no treatment. A preliminary report of the L-PAM study, in which the drug was given for two years, with median follow-up of nine months after initiation of the study, revealed statistically significant delay in the appearance of recrudescent carcinoma in premenopausal women ($p = 0.02$) (39). In the trial comparing 12 monthly cycles of CMF versus no treatment following mastectomy, there were 5.3% relapses in the CMF group versus 24% in the control group ($p < 0.000001$), with observations extending to 27 months, and a mean follow-up period of 14 months. CMF was significantly superior for patients in every subset as classified by age, menopausal status, number of positive axillary lymph nodes, and extent of mastectomy (40). Subsequent analysis reveals significant differences in mortality rates after 32 months. Eleven of 179 patients in the control group have died, whereas only four out of 207 patients treated with CMF have died ($p = 0.03$).

Cooper had initiated treatment of 100 women with four or more positive lymph nodes with CMFVP for nine months following radical mastectomy beginning in 1969. Of these, 73 received chemotherapy alone, and 27 started chemotherapy after conventional radiotherapy. By life-table analysis, 70% of the 73 patients treated with chemotherapy alone are alive without evidence of disease at seven years. The group who received radiotherapy before chemotherapy did not do as well as the group treated with chemotherapy alone (R. G. Cooper, J. F. Holland, O. Glidewell, unpublished data). These data are in dramatic contrast to historical controls—among 14,249 women with one or more positive lymph nodes, 58% were dead in eight years (34). The data demonstrate that the effect of chemotherapy is not solely one of delay, but rather is indicative of a curative impact in some women.

COLORECTAL CARCINOMA

There are approximately 100,000 new cases of colorectal carcinoma in the United States annually, and the overall five-year survival is only about 30% (41). Surgical treatment of this tumor has failed to provide a significant improvement in the survival of operable and therefore potentially curable patients during the past two decades. A technique for intraluminal administration of 5-FU at the time of surgery followed by intravenous 5-FU on the first and second postoperative days has been described, without increased survival among patients with disease limited to the bowel wall (42). Furthermore, a randomized clinical trial employing intraluminal 5-FU, or intraluminal saline, at the time of surgery followed by intravenous 5-FU on the first two postoperative days and five subsequent courses of oral 5-FU over a one-year period, revealed no differences in relapse rate or survival at five years in patients with nodal metastases (43).

A controlled trial employing FUdR on the first three postoperative days followed by a second course approximately six weeks later showed no differences in survival between patients receiving the drug and those treated by surgery alone (44). The Veteran's Administration Surgical Adjuvant Group has reported a study in which patients with curative resection were randomized to receive 5-FU 12 mg/kg intravenously for five successive days beginning 14 days after operation and a second course six weeks later. At 4.5 years, 56% of the 5-FU–treated patients were alive, in contrast to 50% in the control group (45). Although not significant per se, it is consistent with other studies and subset analyses performed on V.A. patients, all of which show a slight trend favoring 5-FU (G. Higgins, personal communication). The current V.A. study entails randomization between an untreated arm and chemotherapy with 5-FU, plus methyl cyclohexylchloroethyl nitrosurea (MeCCNU) in courses repeated at approximately seven weekly intervals for one year. The Central Oncology Group reported a prospective randomized study in which patients received a loading course of intravenous 5-FU within one month of surgical resection, followed by weekly 5-FU maintenance therapy for one year. Preliminary analysis with follow-up times ranging from six months to three years reveals a similar trend in favor of 5-FU that is not yet significantly different (46).

An early cooperative clinical trial involved treatment with thiotepa given at the time of operation into a branch of the portal vein, as well as intraperitoneally and subsequently intravenously on the first and second postoperative days. Because of a high rate of complication, the study was shortly changed, lowering the dosage of thiotepa, eliminating the intraportal dose, and initiating a double-blind randomization. However, among patients receiving the original higher dose of thiotepa, there was a statistically significant increase in seven-year survival in women over the age of 55. When the smaller dose of thiotepa was given, there were no differences between control and treated groups (47).

Recently, a nonrandomized study has been reported, in which patients with colorectal carcinoma penetrating the muscularis with or without node metastases were treated with 5-FU. The drug was given in two five-day courses, the first, four to six weeks after curative surgery and the second, four weeks later. In patients without node metastases, 81.6% of treated patients were free of disease at five years, compared to 58.5% of a group of historical control patients from 1960 to 1965. In those with node metastases, 75% of patients were without evidence of disease at three years, as opposed to 34.3% in a similarly chosen control group (48). These data are so much better than other reports using more 5-FU over a longer period of time that they require confirmation before acceptance. Perhaps the appropriate control for scientific and ethical purposes would be the best 5-FU technique of V.A. studies.

In another nonrandomized study, patients with carcinoma in the muscularis or metastasized to lymph nodes, or those patients believed to have residual tumor in lymph nodes or adjacent organs, and those with peritoneal metastases were treated with 5-FU. A loading course three to four weeks postoperatively together with monthly courses for six to nine months was followed by second-look operations. Among 114 operations in 102 patients, there were eight who had no tumor present and five who were rendered tumor-free by one or more laparotomies. It was suggested that adjuvant chemotherapy followed by a second-look operation can be employed with acceptable morbidity and mortality, and that this approach is of value in defining a cancer-free state or converting patients to it (49).

Immunotherapy has recently been employed as a component of surgical adjuvant treatment. Eighty-three patients with colon cancer spread to regional nodes received treatment with BCG alone or in combination with oral 5-FU, with follow-up times to 30 months. BCG was administered by scarification weekly for three months and then every other week. 5-fluorouracil was given orally as a loading course and then repeated every four weeks for two years. Treatment was started within eight weeks of surgery in 77 patients. When compared to an historical control group, there was prolongation of disease-free interval and overall survival among patients receiving chemoimmunotherapy as well as in patients receiving BCG alone (50).

An ongoing study by the Gastrointestinal Tumor Study Group of patients with colon carcinoma in the muscularis or metastasized to nodes, who have undergone curative resection, involves randomization to receive no treatment, chemotherapy with 5-FU and MeCCNU, immunotherapy with the methanol extraction residue (MER) of BCG, or chemoimmunotherapy. Patients with the same disease stages of rectal carcinoma will be randomized to observation, postoperative radiotherapy, chemotherapy with 5-FU and MeCCNU, or combined radiotherapy and chemotherapy.

BRONCHOGENIC CARCINOMA

The overall five-year survival rates for all patients with bronchogenic carcinoma is approximately 5% (51). Among patients with the least tumor burden according to the TNM classification, i.e. T1, NO, MO (a tumor less than 3 cm in diameter without nodal or other metastases), representing approximately 7% of all lung cancer patients, the survival at 18 months is only 65% (51). In a series of 202 patients dying from whatever cause within one month after curative resection for bronchogenic carcinoma, local or distant metastatic disease was found at autopsy in 35% of patients, implying a much higher frequency of micrometastatic disease (52).

Surgical adjuvant trials employing chemotherapy or chemotherapy plus radiotherapy have not yet shown significant efficacy. The Veteran's Administration Surgical Adjuvant Group conducted four prospective, randomized trials to evaluate single-drug adjuvant chemotherapy after resection of bronchial carcinoma. In 2348 patients involved in trials employing mechlorethamine, cyclophosphamide, or methotrexate given intrapleurally and/or intravenously in courses ranging from three days to 18 months, the five- and ten-year survival rates, as computed by life-table analysis for the treated and control patients in these trials, were found to be essentially the same (53). Four other studies employing short-term postoperative treatment with either mechlorethamine or vinblastine were likewise unsuccessful (54–57). The Swiss Group for Clinical Cancer Research carried out a study involving 189 patients who underwent radical curative resection for bronchogenic carcinoma; following this, patients received either no further therapy or six courses of intravenous cyclophosphamide over a period of two years. After follow-up periods ranging from 3.5 to 5.5 years, 68% of treated patients developed recurrent disease, as opposed to 49% of the control group. Likewise, 63% of the treatment group had died, versus 39% in the control group (58). In two other trials employing cyclophosphamide, increased survival was apparent only for patients with small-cell carcinoma at three years in one of the studies; at five years there were no differences.

None of these differences was statistically significant, and in patients with carcinomas of other cell types, neither harm nor benefit was apparent (55, 59). In a series of 22 patients with Stage I disease treated with 13 courses of cyclophosphamide, 5-FU, and vinblastine over three years following surgery with curative intent, there was increased survival compared to randomized and historical controls up to 32 months after surgery (60). There are also reports of 23 prospective, randomized studies, including 2036 patients with bronchogenic carcinoma employing combined chemotherapy and radiotherapy. The superiority of combination therapy was apparent in only three studies. In no study did the mean or median survival exceed one year (61).

Several studies have been carried out employing immunotherapeutic agents. Life-table analysis of a small study of surgically resected lung cancer patients with Stages I–III disease predicts prolongation of disease-free intervals and increased survival for patients treated with three monthly cycles of high-dose methotrexate with citrovorum rescue plus intracutaneous immunization with specific soluble lung cancer antigen in Freund's complete adjuvant (62). In another nonrandomized study, patients with node metastases underwent radical resection and received postoperative adjuvant immunotherapy with fresh autologous tumor. The median survival of the treated group was 24.5 months, versus 6 months in the control group ($p < 0.01$) (63).

GASTRIC CARCINOMA

Between 75 and 90% of patients in whom gastric cancer is diagnosed are deemed suitable for laparotomy, and 50–60% of these are subsequently eligible for potentially curative resection. Nevertheless, the overall five-year survival rate ranges between 5% and 15% and has not changed significantly in the past 25 years (64). Evaluation of the studies of surgical adjuvant chemotherapy is difficult, as stratification generally included only two broad categories, curative and palliative resections. Rarely was stage, histologic grade, or disease type considered to insure that these characteristics were equally distributed among study arms. Definitions of the surgical procedures employed have not been standardized.

Two early studies employing thiotepa failed to show prolongation of survival for treated patients versus controls (65, 66). A randomized, controlled trial employing two courses of chemotherapy with FUdR or observation revealed comparable survival rates at three years for both groups (67). In three Japanese studies (two intravenous and intraperitoneal regimens of mitomycin C or thiotepa, and a combination of cyclophosphamide and chromomycin A3), there are discrepancies in the results, although there is an overall survival trend favoring chemotherapy (68). A controlled, randomized Russian study is reported, in which patients underwent either radical operation alone or operation plus postoperative chemotherapy with 5-FU given in 2 courses within 3 months of surgery. During the first 18 months after surgery there was a trend, albeit not significant, favoring patients treated with 5-FU, although survival data are not available (69). Nokojima et al report a study in which patients received intravenous mitomycin C either daily or twice a week for 5 weeks immediately following radical surgery. The overall 5-year survival rate significantly favored the treated group: 59.4% of this group was alive, versus 41.5% of the control group (70). The Viennese Working Group reported a pilot study in which patients were given high-dose intermittent chemotherapy with autologous bone marrow transfusion available if necessary. Two to three intravenous infusions of cyclophosphamide at a dose of 20–25 mg/kg were given at intervals of 3–5 days starting 3–6 days after surgery. Cyclophosphamide was continued for 3 years in courses of 400 mg per day given intravenously for 10 days every 4–6 weeks. Six patients received bone marrow reinfusion. Data were analyzed only for patients with Stage III disease. Life-table analysis revealed 5 out of 13 patients alive at 60 months after adjuvant chemotherapy, compared to just under 20% of 166 nonrandomized historical controls treated by surgery alone (71).

A randomized study in which patients with resectable stomach cancer in the antral region received preoperative intraarterial infusions with methotrexate, vinblastine, and mitomycin C or methotrexate, vinblastine, and 5-FU revealed no significant differences in 5-year survival between the treated and control groups (72).

GYNECOLOGICAL CARCINOMA

In endometrial carcinoma, a single adjuvant study has been reported in which patients with Stage I disease (carcinoma confined to the corpus) were treated with intracavitary radium and surgery, or surgery alone, together with Depo-Provera or placebo before hysterectomy and then weekly for 14 weeks. Survival analysis at 4 years indicated no significant improvement with the administration of Depo-Provera. At 4 years, 87% of treated patients were free of disease, versus 92% of those who received a placebo (73).

No major impact in the overall survival in ovarian carcinoma has been made in the past 30 years (74). The 5-year survival rate for patients with Stages I and II disease are only 61% and 33%, respectively (Stage I: growth limited to one or both ovaries with or without ascites; Stage II: growth involving one or both ovaries with extension and/or metastasis to the uterus, fallopian tubes, or other pelvic tissue) (75). In a nonrandom trial, 26 patients with Stages I and IIA disease were treated with postoperative chlorambucil with follow-up ranging from 3½–8 years. There was 1 out of 17 recurrences in the Stage I group and 2 out of 6 in the Stage IIA group. No definitive conclusions can be drawn from these numbers (76).

The Gynecologic Oncology Group has reported preliminary results of a randomized trial for patients with Stage IA and Stage IB ovarian carcinoma who have had definitive surgical treatment with a total abdominal hysterectomy and bilateral salpingo-oophorectomy with optional omentectomy. Patients were then randomized to receive either no further treatment, postoperative radiotherapy, or melphalan. Preliminary analysis revealed respective recurrence rates of 19%, 13.2%, and 0%, and mortality rates of 14.3%, 6.6%, and 0% (77).

A prospective, randomized study, not stratified by stage, involved definitive surgery followed by randomization to either whole-abdominal irradiation by the moving-strip technique with a boost to the pelvis, or to 12 cycles of melphalan.

Melphalan was given in 5-day cycles repeated every 4 weeks pending recovery from hematologic toxicity. All melphalan-treated patients underwent a "second-look" operation following completion of chemotherapy. A year after the beginning of treatment, 85% of the patients with Stage I carcinoma treated with irradiation were without evidence of disease compared to 90% of those who received chemotherapy as well. For Stage II disease, 55% of the irradiation-treated patients and 50% of those given chemotherapy were without evidence of disease. Only one patient who received chemotherapy had a tumor recurrence after 24 months (78). Among a group of 494 patients with advanced ovarian carcinoma treated with melphalan, 103 patients had "second-look" surgery. Sixty-five percent of these survived an additional two years and 34% survived 5 years. Twenty-three had no evidence of disease and chemotherapy was stopped. Subsequently, 5 have had a recurrence of cancer (79). Another retrospective analysis involved patients with Stage II and III disease who, following maximal surgery and postoperative radiotherapy, received chlorambucil or cyclophosphamide concurrently with radiotherapy in 25 patients, and because of interruption of therapy secondary to leukopenia, for 1–4 weeks longer in 8 patients. Among patients who had more than half the tumor excised and no palpable disease postoperatively and who subsequently received chemotherapy, the median survival time was 64+ months versus 26 months for those receiving radiotherapy alone. Among patients with residual masses palpable on abdominal examination postoperatively, there was also a trend toward more complete responses of greater duration among patients receiving chemotherapy in addition to radiation (80).

CONCLUSION

The evidence that the principles of surgical adjuvant chemotherapy developed in experimental animal systems also apply to a variety of neoplastic diseases in man has been clearly demonstrated. Micrometastatic disease can be eradicated with effective chemotherapy in several diseases. Prolongation of disease-free interval, if not cure, is now possible in diseases in which curative surgery alone or in combination with radiotherapy does not achieve these goals. The previously fatal childhood solid tumors—Wilms', Ewing's sarcoma, embryonal rhabdomyosarcoma—are curable in a high percentage of patients appropriately treated with combinations of surgery, radiotherapy, and chemotherapy. The prolongation of the disease-free interval in osteogenic sarcoma has permitted consideration of entirely new surgical approaches for this tumor in which radical amputation has traditionally been employed. The spectacular results achieved in the treatment of Stage II breast cancer may potentially save hundreds of thousands of lives in the coming decade.

Clinically recognizable metastatic disease is rarely curable by any currently available treatment modality. The prolongation of disease-free intervals and production of cures when surgical adjuvant chemotherapy is employed may be partly explained by relatively more circulation, and thus drug delivery to each tumor cell, more favorable cellular kinetics, and a healthier and more immunocompetent host who is better able to withstand drug effects on normal tissues, and to participate in tumor destruction.

Cures of certain patients with neoplastic diseases using surgical adjuvant chemotherapy has increased the incentive to learn more about new and old drugs and their effective use alone and in combination. Chemotherapy, in appropriate combinations with surgery, radiotherapy, and immunotherapy, may well be more efficacious in many clinical situations than the traditional use of single-modality treatment. The data presented in this paper relate solid evidence that the possibility of cure in a variety of neoplastic diseases is real.

Literature Cited

1. Shapiro, D. M., Fugmann, R. A. 1967. A role for chemotherapy as an adjunct to surgery. *Cancer Res.* 17:1098–1101
2. Martin, D. S., Fugmann, R. A. 1960. Clinical implications of the interrelationship of tumor size and chemotherapeutic response. *Ann. Surg.* 151:97–100
3. Schabel, F. M. 1975. Concepts for systemic treatment of micrometastases. *Cancer* 35:15–24
4. Wolff, J. A. 1975. Advances in the treatment of Wilms' tumor. *Cancer* 35: 901–4
5. Everson, R. B., Fraumeni, J. F. Jr. 1975. Declining mortality and improving survival from Wilms' tumor. *Med. Pediatr. Oncol.* 1:3–10
6. Lattimer, J. K., Conway, G. F. 1968. The place of surgery in Wilms' tumor. *J. Am. Med. Assoc.* 204:987–88
7. Gross, R. E., Neuhauser, E. B. D. 1950. Treatment of mixed tumors of the kidneys in childhood. *Pediatrics* 6:843–52
8. Farber, S. 1966. Chemotherapy in the treatment of leukemia and Wilms' tumor. *J. Am. Med. Assoc.* 198:826–36
9. Sutow, W. W., Thomas, W. G., Windmiller, J. 1963. Vincristine (leurocristine) sulfate in the treatment of children with metastatic Wilms' tumor. *Pediatrics* 32:880–87
10. D'Angio, G., Beckwith, J. B., Bishop, H., Breslow, N., Feigl, P., Evans, A., Goodwin, W., Pickett, L., Sinks, L., Sutow, W., Tefft, M., Wolff, J. 1974. The national Wilms' tumor Study—preliminary results. *Proc. Am. Assoc. Cancer Res.* 15:68 (Abstr.)
11. National Wilms' Tumor Study Protocol. 1974. 2:1
12. Dallan, D. C., Coventry, M. B., Scanlon, P. W. 1961. Ewing's sarcoma—A critical analysis of 165 cases. *J. Bone Joint Surg.* 43(A):185–92
13. Boyer, C. W. Jr., Brickner, T. J. Jr., Perry, R. H. 1967. Ewing's sarcoma: Case against surgery. *Cancer* 20:1602–6
14. Pinkel, D. 1962. Cyclophosphamide in children with cancer. *Cancer* 15:42–49
15. Fernandez, C. H., Lindburgh, R. D., Sutow, W. W., Samuels, M. L. 1974. Localized Ewing's sarcoma—Treatment and results. *Cancer* 34:143–48
16. Rosen, G., Wollner, N., Tan, C., Wu, S. J., Majdu, S. I., Cham., W., D'Angio, G. J., Murphy, M. L. 1974. Disease-free survival in children with Ewing's sarcoma treated with radiation therapy and adjuvant 4-drug sequential chemotherapy. *Cancer* 33:384–93
17. Pomeroy, T. C., Johnson, R. E. 1975. Combined modality therapy for Ewing's sarcoma. *Cancer* 35:36–47
18. Ghavimi, F., Exelby, P., D'Angio, G. J., Cham, W., Lieberman, P. H., Tan, C., Mike, V., Murphy, M. L. 1975. Multidisciplinary treatment of embryonal rhabdomyosarcoma in children. *Cancer* 35:677–86
19. Wilbur, J. R., Sutow, W. W., Sullivan, M. P., Castro, J. R., Taylor, H. G. 1971. Successful treatment of rhabdomyosarcoma with combination chemotherapy and radiotherapy. *Proc. Am. Soc. Clin. Oncol.* 7:56 (Abstr.)
20. Selawry, O. S., Hannian, J. 1963. Vincristine treatment of cancer in children. *J. Am. Med. Assoc.* 183:741–46
21. Johnson, D. G. 1975. Trends in surgery for childhood rhabdomyosarcoma. *Cancer* 35:916–20
22. Marcove, R. C., Mike, V., Hajek, J. V., Levin, A. G., Hutter, R. V. P. 1970. Osteogenic sarcoma under the age of 21 —A review of 145 operative cases. *J. Bone Joint Surg.* 52:411–23
23. Dallin, D. C., Coventry, M. B. 1967. Osteogenic sarcoma: A study of 600 cases. *J. Bone Joint Surg.* 49:101–10
24. Marcove, R. C., Mike, V., Hajek, J. V., Levin, A. G., Hutter, R. V. P. 1971. Osteogenic sarcoma in childhood. *NY State J. Med.* 71:855–59
25. Cortes, E. P., Holland, J. F., Wang, J. J., Sinks, L. F. 1972. Doxorubicin in

disseminated osteosarcoma. *J. Am. Med. Assoc.* 221:1132–38

26. Jaffe, N., Pied, D. 1972. Recent advances in the chemotherapy of metastatic osteogenic sarcoma. *Cancer* 30:1627–31
27. Cortes, E. P., Holland, J. F., Wang, J. J., Sinks, L. F., Blom, J., Senn, H., Bank, A., Glidewell, O. 1974. Amputation and adriamycin in primary osteosarcoma. *N. Engl. J. Med.* 291:998–1000
28. Jaffe, N., Frei, E. III, Traggis, G., Bishop, Y. 1974. Adjuvant methotrexate and citrovorum factor treatment of osteogenic sarcoma. *N. Engl. J. Med.* 291:994–97
29. Sutow, W. W., Sullivan, M. P., Fernbach, D. J., Cangir, A., George, S. L. 1975. Adjuvant chemotherapy in primary treatment of osteogenic sarcoma. A Southwest Oncology Group Study. *Cancer* 36:1598–1602
30. Rosen, G., Tan, C., Exelby, P., Beattie, E. J. Jr., Murphy, M. L. 1974. Vincristine (VCR), high dose methotrexate (HDMTX) with citrovorum factor rescue, cyclophosphamide (CY) and adriamycin (ADR) cyclic therapy following surgery in childhood osteogenic sarcoma. *Proc. Am. Assoc. Cancer Res.* 15:1972 (Abstr.)
31. Pratt, C. B., Hustu, H. O., Shanks, E. 1974. Cyclic multiple drug adjuvant chemotherapy for osteosarcoma. *Proc. Am. Assoc. Cancer Res.* 15:19 (Abstr.)
32. Wilbur, J. R., Etcubas, E., Long, T., Glatstein, E., Leavitt, T. 1974. Four-drug therapy and irradiation in primary and osteogenic sarcoma. *Proc. Am. Assoc. Cancer Res.* 15:118 (Abstr.)
33. Rosen, G., Murphy, M. L., Huvos, A. G., Gutierrez, M., Marcove, R. C. 1976. Chemotherapy, *en bloc* resection and prosthetic bone replacement in the treatment of osteogenic sarcoma. *Cancer* 37:1–11
34. Axtell, L. M., Cutler, S. J., Myers, M. H. 1972. *End Results in Cancer Reports No. 4, DHEW Publication # (NIH) 73–272, p. 101.* Bethesda: Natl. Cancer Inst.
35. Fisher, B., Slack, N., Katrych, D., Wolmark, N. 1975. Ten year follow-up results of patients with carcinoma of the breast in a cooperative trial evaluating surgical adjuvant chemotherapy. *Surg. Gynecol. Obstet.* 140:528–34
36. Fisher, B., Ravdin, R. G., Ausman, R. K., Slack, N. J., Moore, G. E., Noer, R. J. 1968. Surgical adjuvant chemotherapy in cancer of the breast: Results of a decade of cooperative investigation. *Ann. Surg.* 168:337–56
37. Leone, L., Rege, V. 1973. Treatment of metastatic, recurrent or inoperable carcinoma of breast with VCR/Pred/5-FU/MTX/Cyclo (Reg. 1) vs. VCR/Pred/5-FU (Reg. II). *Proc. Am. Assoc. Cancer Res.* 14:125 (Abstr.)
38. Canellos, G. P., Pocock, S. G., Taylor, S. III. 1976. Combination chemotherapy for metastatic breast carcinoma: Prospective comparison of multiple drug therapy with L-phenylalanine mustard. *Cancer.* 38:1882
39. Fisher, B., Carbone, P., Economou, S. G., Frelick, R., Glass, A., Lerner, H., Redmond, C., Zelen, M., Band, P., Katrych, D. L., Wolmark, N. 1975. L-phenylalanine mustard (L-PAM) in the management of primary breast cancer. A report of early findings. *N. Engl. J. Med.* 292:117–22
40. Bonadonna, G., Brusamolino, E., Valagussa, P., Rossi, I., Brugnatelli, L., Brambilla, C., DeLena, M., Tancini, G., Bajetta, E., Musumeci, R., Veronesi, U. 1976. Combination chemotherapy as an adjuvant treatment in operable breast cancer. *N. Engl. J. Med.* 294:405–10
41. Silverberg, E., Holleb, A. I. 1975. Major trends in cancer: Twenty-five year survey. *Cancer* 25(1):2–8
42. Rousselot, L. M., Cole, D. R., Grossi, C. E., Conte, A. J., Gonzalez, E. M., Pasternak, B. S. 1972. Adjuvant chemotherapy with 5-FU in surgery for colorectal cancer: Eight year progress report. *Dis. Colon Rectum* 15:174–79
43. Lawrence, W., Terz, J. J., Horsley, S. III, Donaldson, M., Lovett, W. L., Brown, P. W., Ruffner, B. W., Regelson, W. 1975. Chemotherapy as an adjuvant to surgery for colo-rectal cancer. *Ann. Surg.* 181:616–23
44. Dwight, R. W., Humphrey, E. W., Higgins, G. A., Keehn, R. J. 1973. FUdR as an adjuvant to surgery in cancer of the large bowel. *J. Surg. Oncol.* 5:243–50
45. Higgins, G. A., Dwight, R. W., Smith, J. V., Keehn, R. J. 1971. Fluorouracil as an adjuvant to surgery in carcinoma of the colon. *Arch. Surg.* 102:339–43
46. Grage, T., Cornell, G., Strawitz, J., Jonas, K., Frelick, R., Metter, G. 1975. Adjuvant therapy with 5-FU after surgical resection of colo-rectal cancer. *Proc. Am. Assoc. Cancer Res.* 15:258 (Abstr.)

47. Holden, W. D., Dixon, W. J., Kuzman, J. W. 1967. The use of triethylene thiophosphoramide as an adjuvant to the surgical treatment of colo-rectal carcinoma. *Ann. Surg.* 165–481
48. Li, M. C., Ross, S. T. 1976. Chemoprophylaxis in patients with colo-rectal cancer. Prospective study with 5 year follow-up. *J. Am. Med. Assoc.* 235:2825–28
49. Mackman, S., Ansfield, S. J., Ramirez, G., Curreri, A. R. 1974. A second look at the second look operation of colonic cancer after the administration of fluorouracil. *Surgery* 128:763–66
50. Mavligit, G. M., Burgess, M. A., Seibert, G. B., Jubert, A. V., McBride, C. M., Gehan, E. A., Gutterman, J. U., Khankhanian, N., Speer, J. F., Martin, R. C., Copeland, E. M., Hersh, E. M. 1976. Prolongation of postoperative disease-free interval and survival in human colo-rectal cancer by BCG or BCG plus 5-fluorouracil. *Lancet* 1:871–76
51. Mountain, C. F. 1973. Keynote address on surgery in the therapy for lung cancer: Surgical prospects and priorities for clinical research. *Cancer Chemother. Rep.* (Pt. 3) 4:19–24
52. Matthews, M. J., Kanhouwa, F., Pickren, J., Robinette, D. 1973. Frequency of residual and metastatic tumor in patients undergoing curative surgical resections for lung cancer. *Cancer Chemother. Rep.* (Pt. 3) 4:63–67
53. Shields, T. W., Robinette, C. D., Keehn, R. J. 1974. Bronchial carcinoma treated by adjuvant cancer chemotherapy. *Arch. Surg.* 109:329–33
54. Curreria, A. 1962. Nitrogen mustard as an adjuvant to pulmonary resection in the treatment of carcinoma of the lung. *Cancer Chemother. Rep.* 16:123–24
55. Higgins, G. 1972. The use of chemotherapy as an adjuvant to surgery for bronchogenic carcinoma. *Cancer* 30: 1383–87
56. Flack, N. J. 1970. Bronchogenic carcinoma. Nitrogen mustard as a surgical adjuvant and factors influencing survival. *Cancer* 25:987–1002
57. Crosbie, W. A., Camdar, H. H., Belcher, J. R. 1966. A control trial of vinblastine sulfate in the treatment of carcinoma of the lung. *Br. J. Dis. Chest* 60:28–35
58. Brunner, K. W., Marthaler, T. H., Mueller, W. 1973. Effects of long-term adjuvant chemotherapy with cyclophosphamide (NSC-29271) for radically resected bronchogenic carcinoma. *Cancer Chemother. Rep.* (Pt. 3) 4: 125–32
59. Miller, A. B. 1972. (A Medical Research Council Working Party) Study of cytotoxic chemotherapy as an adjuvant to surgery in carcinoma of the bronchus. *Br. Med. J.* 2:421–28
60. Karrer, K., Pridun, N., Zwintz, E. 1973. Chemotherapeutic studies in bronchogenic carcinoma by the Austrian Study Group. *Cancer Chemother. Rep.* (Part 3) 4:207–13
61. Selawry, O. S. 1974. The role of chemotherapy in the treatment of lung cancer. *Semin. Oncol.* 1:259–72
62. Stewart, T. H. M., Hollinshead, A. C., Harris, J. A. 1976. Immunochemotherapy (IC) of lung cancer. *Proc. Am. Soc. Clin. Oncol.* 16:305 (Abstr.)
63. Takita, H., Minowada, J., Han, T., Takada, M. 1975. Adjuvant immunotherapy with autologous tumor in human lung cancer. *Int. Conf. Immunobiol. Cancer, 24th* (Abstr.)
64. Moertel, C. G. 1973. In *Cancer Medicine,* ed. J. F. Holland, E. Frei III, pp. 1532–33. Philadelphia: Lea & Febiger
65. V.A. Cooperative Surgical Adjuvant Study Group. 1965. Use of thiotepa as an adjuvant to the surgical management of carcinoma of the stomach. *Cancer* 18:291–97
66. Dixon, W. J., Longmire, W. P., Holden, W. D. 1971. Use of triethylenethiophosphoramide as an adjuvant to the surgical treatment of gastric and colorectal carcinoma: 10-year followup. *Ann. Surg.* 173:26–39
67. Serlin, O., Wolkoff, J. S., Amadeo, J. M., Keehn, R. J. 1969. Use of 5-fluorodeoxyuridine (FUdR) as an adjuvant surgical management of carcinoma of the stomach. *Cancer* 24:223–28
68. Koyama, Y., Kimura, T., Takemasa, Y. 1970. Chemotherapy as an adjuvant to surgery in stomach cancer. *Prog. Antimicrob. Anticancer Chemother.* 2:242–49
69. Blokhina, N. G., Garin, A. M., Moroz, L. V., Pershin, M. P., Chumakina, S. I., Gunka, I. I., Pantophel, A. M., Steklenev, N. A., Zirin, M. A., Grjaznova, I. N., Shain, A. A., Pecherskaja, B. G., Saveliev, N. P., Bashirova, N. G., Kuzmin, V. P., Budarina, E. M., Svedencov, E. P., Phateeva, K. V., Iunusmetov, I. R., Silitrin, N. P., Shklovskii, G. S., Balas, A. N., Ushivceva, A. E. 1972. Treatment with 5-fluorouracil in prophylaxis of relapses and metastases of stomach cancer. *Neoplasma* 19:351–56

70. Nakajima, T., Kayatani, T., Fukami, A., and Ohashi, I. 1975. Effect of adjuvant chemotherapy with Mitomycin-C on the recurrence of gastric cancer after radical surgery. *Jpn. J. Cancer Clin.* 21:879–904
71. Karrar, A. 1972. Importance of dose schedules in adjuvant chemotherapy. *Cancer Chemother. Rep.* 56:35–45
72. Fujimoto, S., Akao, T., Itoh, B., Itaru, K., Koyano, K., Kituskawa, Y., Takhashi, M., Minami, T., Ishigami, H., Miyazaki, M., Itoh, K. 1976. A study of survival in patients with stomach cancer treated by a combination of preoperative intra-arterial infusion therapy and surgery. *Cancer* 37:1648–53
73. Lewis, G. C., Slack, N. H., Mortel, R., Bross, I. K. J. 1974. Adjuvant progestogen therapy in the primary definitive treatment of endometrial carcinoma. *Gynecol. Oncol.* 2:368–76
74. DeVita, V. T., Wasserman, T. H., Young, R. C., Carter, S. K. 1976. Prospectives and research in gynecologic oncology. *Cancer* 38:509–25
75. Griffiths, C. T. 1973. See Ref. 64, p. 1713
76. Julian, C. G., Woodruff, J. D. 1969. The role of chemotherapy in the treatment of primary ovarian malignancy. *Obstet. Gynecol. Surv.* 24:1307–42
77. Hreshchyshyn, M. M. 1975. *Natl. Cancer Inst. Monogr. 42,* pp. 155–65
78. Smith, J. P., Rutledge, N., Delclos, L. 1975. Postoperative treatment of early carcinoma of the ovary: A random trial between postoperative irradiation and chemotherapy. *Natl. Cancer Inst. Monogr.* 42:149–53
79. Smith, J. P., Rutledge, N. N. 1975. Chemotherapy in advanced ovarian cancer. *Natl. Cancer Inst. Monogr.* 42:141–43
80. Griffiths, C. T., Grogan, R. H., Hall, T. C. 1972. Advanced ovarian cancer: Primary treatment with surgery, radiotherapy, chemotherapy. *Cancer* 29:1–7

Ann. Rev. Med. 1977. 28:489–515
Copyright © 1977 by Annual Reviews Inc. All rights reserved

BCG AS ADJUVANT IMMUNOTHERAPY FOR NEOPLASIA[1]

◆7257

Evan M. Hersh, M.D., Jordan U. Gutterman, M.D., and Giora M. Mavligit, M.D.
Department of Developmental Therapeutics, The University of Texas System Cancer Center, M. D. Anderson Hospital and Tumor Institute, Houston, Texas 77030

INTRODUCTION AND BACKGROUND

There is a solid scientific basis for immunotherapy with Bacillus Calmette-Guérin (BCG), as well as for other approaches to immunotherapy of cancer in man. This is based on observations in both spontaneous and experimental animal tumors, as well as human cancer. The fundamental basis for immunotherapy is the presence of tumor-associated antigens that the host can recognize as foreign, and immune responses and other host defense mechanisms to these antigens (1). This was first demonstrated in animal models, particularly models in which strong tumor antigens exist, such as those associated with viral and carcinogen induction, although some spontaneous tumors apparently do not express such antigens. During the last decade, evidence has developed that tumor-associated antigens and immune responses, as well as other host defense mechanisms directed at these antigens, also exist in human cancer (2).

A surprisingly varied spectrum of anti-tumor host defense mechanisms has been identified in various experimental tumor models, as well as human tumors (3). These include complement-dependent cytotoxic antibody and other humoral nonantibody factors with selective toxicity for tumor cells. These include macrophage enzymes, tumor necrosis factor (4), and lymphotoxin (5). There are a variety of leukocytes

---

[1] Supported by Grant number CA-05831 and Grant number CA-14984, from the National Cancer Institute, National Institutes of Health, Bethesda, Maryland. Drs. Gutterman and Mavligit are recipients of Public Health Research Career Development Awards #1-K04-CA-71007 and #-K04-CA-00130, respectively, from the National Institutes of Health, Bethesda, Maryland 20014.

that can also kill tumor cells (1). The cellular mechanisms include T-cell–mediated cytotoxicity, K-cell–mediated cytotoxicity, antibody-dependent cell–mediated cytotoxicity, and macrophage-mediated killing of tumor cells (6). There is considerable experimental evidence that the latter is an important mechanism of antitumor host defense and may be one of the major host defense factors activated by BCG (7). There is also a solid clinical basis for the types of tumor immunity and antitumor host defense outlined above. Thus, both humoral and cell–mediated antitumor immunity have been demonstrated in human malignancies, and deficiencies or declines in this immunity are associated with a poor prognosis (8).

Observations on the general immunocompetence of cancer patients are also highly relevant to immunotherapy. Lymphocyte infiltration of primary tumors and histiocyte proliferation in regional lymphoid sinuses are associated with a good prognosis in a number of solid tumors, whereas the absence of such cellular infiltration and proliferation is associated with a worse prognosis (9). This is particularly true in certain diseases such as breast cancer, in which this type of cellular response is more important than other prognostic factors, such as the number of involved positive lymph nodes. Recently, macrophage content of tumors has been investigated. A high macrophage content has good prognostic implications (10). There is a progressive decline in immunocompetence with progressive tumor growth (11). This has been demonstrated using primary delayed hypersensitivity reactions to antigens such as dinitrochlorobenzene (DNCB), recall delayed hypersensitivity reactions to antigens such as candida, as well as a variety of in vitro lymphocyte-function studies, including blastogenesis and MIF production; T- and B-cell populations also decline. Complicating or accentuating this decline in immunocompetence is the immunosuppressive effect of cancer therapy (11). Surgery, radiotherapy, and chemotherapy are all immunosuppressive. The immunosuppressive effects of surgery on cell–mediated immunity last up to a month. The suppressive effects of radiotherapy on circulating T- and B-lymphocyte levels can last for years, while intermittent chemotherapy (the common form of chemotherapy today) usually has transient immunosuppressive effects with rebound between courses of treatment. There is a cumulative irreversible immunosuppressive effect after repeated courses of treatment.

Finally, the well-established relationship between immunocompetence and prognosis is relevant to immunotherapy (12). Patients with primary tumors and patients with disseminated or metastatic malignancy manifest this phenomenon. The inflammatory response is impaired in solid-tumor and lymphoma patients and the degree of impairment is related to prognosis. Patients with primary solid tumors who cannot mount a delayed hypersensitivity response after sensitization with DNCB may be free of disease by surgery, but have early recurrence. Patients who can manifest delayed hypersensitivity have a much better prognosis. If the DNCB response declines after surgery, early recurrence is likely. Patients with lung cancer who have an impaired delayed hypersensitivity response to the recall antigen candida have poor survival, while patients with squamous cell carcinoma of the lung who have mumps responses smaller than 5 mm of induration have a poor response to therapeutic intervention. Patients with breast cancer who have a subnormal in vitro lymphocyte blastogenic response to phytohemagglutinin (PHA) have a poor prognosis relative to patients who have a normal or vigorous PHA response. In acute leukemia, good prognosis could be predicted on the basis of the recall antigen response, the in vitro lymphocyte response to mitogens or antigens, or the ability to develop a delayed hypersensitivity response to the primary protein antigen keyhole limpet hemocyanin (KLH). Low levels of circulating B lymphocytes are associated with a poor prognosis both in primary solid-tumor patients before surgery and in patients with acute leukemia after intensive intermittent chemotherapy. These are just a few examples of the many observations that have been made relating immunocompetence and prognosis.

There are also factors in the serum of cancer patients that block lymphocyte proliferation and cell-mediated killing or inhibition of tumor cell proliferation when tested in various in vitro systems (13). These blocking factors were initially thought to be antibody, but are probably circulating soluble antigen or antigen-antibody complexes (13). For example, in lung cancer we have recently observed a direct correlation between the degree of elevation of circulating immune complexes and diminished survival (14). The multiple factors outlined in this brief introduction form a rationale for immunotherapy and a scientific structure on which the objectives of immunotherapy can be developed. These are briefly discussed in the next section.

OBJECTIVES OF IMMUNOTHERAPY

Based on the above, a general scheme regarding objectives of immunotherapy can be developed. If a tumor-bearing individual's general immunocompetence is impaired, or if the individual is receiving immunosuppressive therapy, attempts can be made to restore immunocompetence to a normal level. A number of immunorestorative agents that increase the size of the T-cell population or improve T-cell function have been described and include levamisole (15), thymosin (16), and transfer factor (17). Microbial adjuvants that include BCG probably also have this as one of their functions. If the tumor-bearing individual has diminished or absent immunological reactivity to the tumor-associated antigens, one can attempt to restore that reactivity. This can be accomplished by adoptive immunotherapy with transfer factor, possibly with "immune" ribonucleic acid (immune RNA) (18) or other products of specifically immune lymphocytes, by passive immunotherapy with specific antisera (19), or by active-specific immunotherapy. The latter refers to immunization with tumor cells or tumor antigen.

If the individual has both general immunocompetence and adequate specific tumor immunity, but is at high risk of recurrent disease or has progressive measurable tumor, one must assume that local-regional host defense factors are deficient, inadequate for the size of the tumor, or blocked by immunosuppressive factors released by the tumor and active in the tumor's local environment. In this case, the objective of immunotherapy would be to activate local defense factors, including lymphocyte release of mediators, local antibody production, and direct lymphocyte and macrophage killing of tumor cells. If circulating blocking factors or local high concentrations of blocking factors are responsible for a failure of effective antitumor host defense, activation of the reticuloendothelial system (RES) to help clear these blocking factors, and local activation of macrophages, would be indicated. Augmented specific tumor immunity and augmented nonspecific antitumor host defense mechanisms might also act in concert with such conventional approaches as chemotherapy, and a combination approach might kill a larger number of tumor cells than either approach alone. Restoration of general immunocompetence, boosting of specific tumor immunity, activation of local and regional host defense mechanisms, and activation of the RES have all been demonstrated as activities associated with the administration of BCG.

From the objectives of immunotherapy outlined above, the major approaches to immunotherapy and their rationale become evident (20). Active nonspecific immunotherapy, which refers to the administration of microbial adjuvants, augments cell-mediated immunity, the antibody response, macrophage activity, and particle clearance by the RES. Active specific immunotherapy, referring to immunization with intact or modified tumor cells or tumor antigen, is designed to specifically increase cell-mediated or humoral antitumor immunity. Adoptive immunotherapy refers to the transfer, from a specifically immune host to a non- or hypoimmune recipient, of the specific immune reactivity of the donor. This can be carried out with intact lymphoid cells or cellular extracts, such as immune RNA or transfer factor. In animal models, effective immunotherapy with cell transfer usually requires syngeneic immune cells (21). This would be impractical in the human setting. Recently, the concept of adoptive immunotherapy has been broadened to include the approach of transferring general immunocompetence from a competent donor to an incompetent recipient. Thymosin and possibly transfer factor fall into this category. Passive immunotherapy, referring to the transfer of specific antitumor antibody from immune donor to nonimmune recipient, has a certain potential, but has not been intensively investigated (22). The nature of the antibody could be complement-dependent, cytotoxic, deblocking, or associated with antibody-dependent, cell-mediated cytotoxicity. It could also function with macrophages as cytophilic or opsonizing antibody. Local and regional immunotherapy refers to the administration of an active nonspecific immunotherapeutic microbial reagent or some other immunotherapeutic agent directly into a tumor, or into the regional lymphatic drainage of a primary or secondary tumor after its removal (23). The resultant local inflammatory and delayed hypersensitivity response can kill tumor by means of a bystander effect through the release of lymphokines or through the direct effect of activated macrophages. As is described below, the direct tumor cell-adjuvant-host defense cell interaction often also results in augmentation of systemic cell-mediated immunity. The final approach to immunotherapy that should be discussed is the use of the immunorestorative agent levamisole. This is the prototype of a new class of chemical restorative agents that increase T-cell number and function by increasing cellular maturation and expression of cell-surface markers through presumed activity on the cyclic nucleotide system (24). Finally, combinations of the above approaches, particularly of active nonspecific immunotherapy, with one of more of the others, have been shown to be effective and should be seriously considered for more extensive study in man.

BACKGROUND DATA ON BCG ACTIVITY

BCG is the abbreviation for Bacillus Calmette-Guérin. This is an attenuated strain of *Mycobacterium bovis*, which was developed from a more virulent strain in 1908 by Calmette and Guérin, of the Pasteur Institute (25). All BCG preparations throughout the world are derived from this original strain. Culture techniques vary widely and genetic drift has occurred, so that now the available strains vary in their biological and immunological characteristics. In addition, methods of preparation yield products that vary in the number of viable organisms, the number of dead organisms, the amount of free soluble antigen, and the dispersion of the organisms in terms of whether a "viable" or colony-forming unit consists of one or up to as many as 50 organisms (26). Some of the more commonly available strains or preparations include the Pasteur, Glaxo, Tice, Connaught, and Montreal strains. These are all commercially available for prophylaxis against tuberculosis and for experimental immunotherapy. They are generally provided lyophilized, and are reconstituted with a diluent free of bacteriostatic agents. Viable units vary greatly from preparation to preparation and from lot to lot, but are usually within the range of 1×10^7 to 5×10^8 units per ml. There is variation in the number of viable units because of a variable degree of clumping. In addition to these preparations, considerable experimental and some clinical work is being done with a preparation grown in suspension in a synthetic medium (27). This material has potential advantages because it is fully viable, and contains few dead organisms and essentially no free soluble antigen. Dead organisms and free soluble antigen have been implicated in poor adjuvant activity of some BCG preparations (26). Cutaneous administration of BCG in some, but not all, clinical trials has been shown to provide 30–80% protection against the development of tuberculosis in susceptible populations and is clinically indicated in the appropriate populations (28, 29). The antitumor use of BCG is experimental, carried out under an Intent to use New Drug, and must be done by an approved clinical investigator or under his supervision, with appropriate institutional approval and protection of human rights.

Many studies have demonstrated that BCG has potent immunological effects. These effects include increased reticuloendothelial-system function, namely increased clearance of bacteria such as *Listeria monocytogenes* (30); increased clearance of carbon particles (31); augmentation of the antibody response (32); increased delayed hypersensitivity responses to a variety of antigens (33), including complex antigens such as sheep red blood cells; increased activity of peritoneal macrophages, in terms of metabolic activity, phagocytic activity, and ability to kill tumor cells (34); and increased resistance to infection, even in the face of leukopenia (35). BCG also has the capacity to increase the resistance to immunological paralysis. For example, in mice given a large dose of sheep red blood cells, which would usually result in the development of antibody and the failure of development of delayed-type hypersensitivity, if BCG is administered prior to the antigen, delayed hypersensitivity develops because of RES activation and improved clearance of antigen-antibody complexes (33). It is likely that several of these mechanisms may be involved in the antitumor activity of BCG.

In addition, and of potential relevance to its antitumor activity, it has been recently documented by several investigators, using antigen binding and some immunoelectron microscopy, that BCG cross-reacts with the antigens on the surface of some tumor cells (36). Thus, BCG immunotherapy may operate via an increase in cross-reactive antitumor immunity. Recently, a specific increase in cell-mediated tumor immunity has been observed after BCG immunization in an animal model. Thus, BCG immunotherapy may not only be active nonspecific, but may also have an active specific component.

The immunological effects of BCG are both strain- and preparation-dependent (26). An increased proportion of dead organisms and free soluble antigen diminishes the ability of BCG to stimulate cell-mediated immunity, and augments the antibody response. Different strains augment cell-mediated immunity to different degrees. In general, the more virulent and rapidly growing the strain, the more potent the stimulator of cell-mediated immunity. Recently, in a transplantable, carcinogen-induced guinea pig hepatoma system described more completely below, strain differences in immunotherapeutic efficacy were also demonstrated. Fully viable Pasteur BCG, used in scarification after surgical extirpation of tumor, delayed tumor recurrence more effectively than lyophilized Pasteur-strain or fully viable Tice-strain BCG (37). The extent to which this will be relevant for human immunotherapy is unknown at present, but it must certainly be considered in the design and evaluation of clinical trials.

The immunological effects of BCG administration have also been studied in man. In our own studies of BCG immunotherapy, BCG administration by scarification converts the tuberculin reaction in approximately 80% of patients. Patients with minimum tumor burden show augmentation of delayed hypersensitivity reactions to recall antigens such as dermatophytin and streptokinase-streptodornase (varidase), and to primary antigens such as KLH, compared to untreated controls (38). However, in a small series we have not observed increased reactivity to DNCB after BCG therapy (39). Certain strains of BCG can significantly increase the circulating lymphocyte response to T-cell mitogens, but apparently have no effect on the lymphocyte response to B-cell mitogens (40). T- and B-cell levels in the peripheral blood are not altered by BCG immunotherapy. Finally, the presumed immunological functions of BCG immunotherapy, namely RES and macrophage activation, have not been studied in man.

Two additional aspects of BCG administration are worthy of mention at this point. Intravenous administration of BCG to rats markedly depresses hepatic microsomal enzyme function and specific activity (41). This may have important implications for a combination of chemotherapy and immunotherapy when the chemotherapeutic agent is either activated or degraded by these enzymes. Also, BCG administration markedly sensitizes mice to subsequent administration of endotoxin. A factor that causes regression of tumor and inhibits tumor-cell proliferation in vitro appears in the serum of animals treated with BCG followed by endotoxin. This material has been demonstrated to be a macrophage enzyme or group of macrophage enzymes with tumor-inhibitory properties that can be demonstrated in vitro. It is not specific, and can kill a number of human melanoma cell lines. This material has been referred to as tumor necrosis factor (4). The extent to which tumor necrosis factor generation is a part of the mechanism of action of BCG is not known at this time.

ADMINISTRATION OF BCG

There are numerous routes for BCG administration. Most commonly, BCG is administered into the skin. Methods of administration into the skin include scarification, multipuncture plate, Heaf gun, intradermal, and subcutaneous administration. In the scarification approach, a grid of scratches is made in the skin with the bevel of an 18-gauge needle. BCG is spread in the serosanguineous exudate on the scarified area, dried with cool air from a hair dryer, and protected from contamination and moisture for 24 hours. In the multipuncture and Heaf gun approaches, BCG is painted on the skin, after which a series of 1-mm needles or tines are driven through the painted area, driving some of the BCG organisms into the dermis. Intradermal and subcutaneous administration are with standard needle and syringe techniques. The dose of BCG administered intradermally or subcutaneously can be accurately quantitated, but the dose administered by the other cutaneous methods cannot. However, animal studies have demonstrated that more organisms are delivered into the skin, and more reach the regional draining lymph nodes by the scarification than by the multipuncture technique (42). BCG administered by scarification or multipuncture technique results in an inflammatory lesion; the shape and the size of the application peaks in about one to two weeks, and is usually healed by two to four weeks. On repeated administration, flaring of previously treated sites is common. Flare reactions usually last one to two weeks. Regional adenopathy is common and both flaring and regional adenopathy correlate with a good prognosis (43) in patients getting BCG by scarification.

In contrast to these usually limited reactions, intradermal and subcutaneous administration results in chronic ulcerated draining lesions, which may last for years, and commonly last for months. These latter routes of administration are also associated with a significant incidence of systemic BCG disease (if a high dose of BCG is used), which does not occur with scarification or multipuncture techniques (44). Administration of BCG by injection into primary or metastatic cutaneous lesions may also be considered a form of cutaneous administration. Its characteristics are similar to that of intradermal or subcutaneous administration. An alternative to this type of local immunotherapy is epilesional scarification, which is not associated with severe toxicity (45).

BCG may also be administered orally (46). This route has been hypothesized to be useful against gastrointestinal malignancies; however, conversion of PPD responses is very slow and systemic reactions are almost absent with this route. In at least one experimental animal model, oral BCG was ineffective in terms of antitumor activity (47). BCG may also be administered intravenously. This has been effective in at least one study of acute leukemia (48). Apparently, low doses are not associated with systemic BCG disease. Intracavitary BCG administration is of increasing interest. In man, this would include intrapleural BCG after pneumonectomy for lung cancer (49), and intraperitoneal administration of BCG after removal of cancer of an intra-abdominal organ (50). These routes of administration would presumably activate the local host defense mechanisms in the regional drainage of the primary tumor. This is of already proven efficacy in lung cancer, as is outlined below.

The usual symptomatic side effects associated with cutaneous administration of BCG have been well characterized (44). After repeated administration, patients usually experience 24–48 hr of low-grade fever, chilly sensations, mild malaise, mild arthralgias and myalgias, and easy fatigability. These usually do not interfere with normal daily activities. Less commonly, patients may experience high fever, shaking shills, severe flu-like syndrome, headache, nausea, and vomiting. Local reactions, including erythema, induration, tenderness, mild pain, and pruritis, are common and desirable, as is regional lymphadenopathy. BCG-induced regional lymphadenopathy may be confused with tumor recurrence in regional lymph nodes and its differential diagnosis is obviously very important. Severe local or systemic side effects may require dosage regulation, which is usually carried out in half-log steps. Untoward complications of BCG immunotherapy is described in a subsequent section.

ANIMAL MODELS OF IMMUNOTHERAPY WITH BCG

Immunotherapy with BCG is firmly based in animal model work. In 1959, based on its ability to increase reticuloendothelial system activity, several investigators observed that the administration of BCG increased the resistance of animals to subsequently transplanted experimental tumors (51, 52). It was subsequently demonstrated that a variety of viral and carcinogen-induced tumors, as well as spontaneous tumors such as AKR leukemia in mice, are delayed in development, reduced in incidence, or prevented by pre-treatment with a single dose of BCG (53). In animals with established tumors, BCG can be curative when given intralesionally (54) or intravenously (55). For example, in guinea pig hepatoma (54), intralesional BCG injection of a 7-day-old primary tumor implant causes local tumor regression, complete inhibition of the growth of metastatic tumor in the regional lymph nodes, and induces strong systemic tumor resistance. The latter is manifested by increased resistance to subsequent tumor-cell challenge. The induced tumor immunity is completely tumor type-specific. This system is relevant to human cancer, since this is a weakly antigenic tumor. In rats with a transplantable fibrosarcoma, intravenous BCG causes regression of established artificial pulmonary metastases, as well as artificially induced pleural metastases (55). When administered systemically, either alone or in combination with a tumor-cell vaccine, BCG can prolong surgery or chemotherapy-induced remissions of experimental tumors in animals (56).

The most effective routes of BCG administration in these models seem to be those that are regional or in close approximation to the primary tumor. The mechanism of action of BCG has been elucidated in the various animal models. While it is active in weakly antigenic tumors, such as guinea pig hepatoma, it is most active in antigenic tumors and, indeed, some weakly antigenic tumors are enhanced by BCG administration (57). The mechanism of action seems to vary depending upon the type of tumor, the status of the tumor, and the species under study. Broadly speaking, three major mechanisms of action have been identified. These include augmentation of classical cell-mediated immunity, activation of macrophages with subsequent macrophage-mediated killing of tumor cells, and, as mentioned above, cell-mediated and/or humoral immunity to antigens on BCG organisms that cross-react with tumor cells. The strong evidence for the role of macrophages comes from direct histological studies of tumor undergoing regression after intralesional therapy with BCG (58), from the fact that BCG-activated macrophages have been shown to selectively kill tumor cells (34), and from the fact that macrophages from animals cured of the guinea pig hepatoma by intralesional BCG kill tumor cells at an effector target-cell ratio of 10:1, whereas lymphocytes from the same animals kill only in a ratio of 10,000:1 (59). Direct macrophage lysosome injection into tumor cells has been observed. This further supports the concept of the macrophage role.

One animal model, the guinea pig hepatoma mentioned above, has been extensively used to study BCG immunotherapy (60, 61). Using this model, optimal conditions for BCG immunotherapy have been established. The tumor must be relatively small, the dose of BCG must be relatively large, and there must be close approximation between the BCG and the tumor; even low burdens of tumor cells at distant sites can abrogate the immunotherapeutic effect, and the strain and viability of the BCG are important. Animal models such as this can be used to ask clinically relevant questions regarding BCG, as well as other forms of immunotherapy.

BCG AND CANCER PREVENTION

An intriguing and as yet unresolved question regarding BCG is whether the incidence of malignant disease is lower in individuals inoculated with BCG as prophylaxis against tuberculosis, and whether BCG administration, particularly in infancy, could be used to prevent the subsequent development of malignant disease, particularly acute leukemia and malignant lymphoma of childhood. In 1970, it was first reported that there was a 50% reduction in the leukemia death rate among neonatally vaccinated children (62). In 1972, workers in Chicago reported an 85% reduction in deaths from leukemia among an ethnically homogeneous group of children inoculated with BCG neonatally (63). This latter study has been recently reanalyzed by the National Cancer Institute, which found a 50% reduction in deaths from leukemia and all cancers in the vaccinated, compared to a control group (64).

In contrast to these encouraging observations, when the cancer mortality and/or incidence was investigated in 3 controlled clinical trials of BCG administration as prophylaxis against tuberculosis, no statistically significant difference in leukemia incidence or other cancer incidence was noted (64). There was some reduction in leukemia mortality, but some increased mortality from malignant lymphoma. It should be noted that none of these studies involved neonatal immunization. However, no difference in leukemia mortality was observed in populations in Canada and Scotland, who did receive neonatal BCG immunization, compared to other parts of the world where this is not done (65). Nor were there differences in mortality from leukemia in Scandinavia, where populations are either vaccinated in the newborn period, or at approximately age 13 (66). These data suggest that the question of leukemia and other cancer mortality, as influenced by neonatal BCG administration, cannot be definitively answered at this time. However, the bulk of evidence suggests no significant effect.

HISTORICAL ASPECTS OF BCG IMMUNOTHERAPY OF CANCER IN MAN

In 1969, Mathé and co-workers reported remission prolongation by BCG immunotherapy in acute lymphoblastic leukemia of childhood (67). The study was based on animal experiments in which he and his co-workers demonstrated that the administration of BCG alone, or BCG plus tumor cells, to mice with L1210 leukemia prolonged survival if the tumor burden was under 10^5 cells (68). Thirty patients with chemotherapy-induced complete remission were allocated to receive either 75 mg of BCG by scarification every 4 days for the first month and then weekly, or the same plus 4×10^7 pooled allogeneic irradiated leukemic cells. Both forms of immunotherapy were equivalent in their effect and, at approximately 3 years, 50% of the 20 immunotherapy patients were still in remission. In contrast, all of the ten controls had relapsed by approximately 120 days. Apparently 5 of these patients have remained free of disease to the present time. The fresh liquid form of Pasteur-strain BCG was used in this study. In 1970, Morton and associates reported the results of intralesional BCG immunotherapy of cutaneous metastatic malignant melanoma (69). Patients with metastatic disease were evaluated for the presence of circulating anti-melanoma antibody and for the ability to be sensitized to DNCB. Their tumor nodules were injected with 0.1 ml of Glaxo-strain BCG. Twenty-five patients were treated. Regression of injected nodules was seen in all 19 who were DNCB-positive, and in none of those who were DNCB-negative. Regression of uninjected nodules was also seen in 4 of the 19 patients. It was the impression that meaningful prolongation of remission and survival was seen in some of these patients.

In 1965, Villasor, working in the Philippines, treated patients with repeated courses of chemotherapy plus repeated BCG vaccination, using 0.2 mg of Pasteur-strain BCG administered intracutaneously (70). The patients had various types of metastatic malignancy including breast, lung, thyroid, prostate, and colon. Apparently subjective and objective responses increased from about 39 to about 74%, and the one-year survivors increased from 35 to 59%, while the two-year survivors increased from 6 to 16%. Questions regarding the controls, the comparability of the groups, the objective or subjective nature of the responses, and the lack of publication in an international journal prevented the results of this study having a major impact.

In 1972, Bluming and co-workers carried out a comparative study of two forms of BCG immunotherapy in African patients following removal of primary melanoma of the foot (71). Six patients received 0.1 ml of Glaxo BCG intradermally every 4th day, and 6 patients received 150 mg of lyophilized Pasteur BCG every 4th day by scarification, the duration of therapy in both groups being 1 month. The group receiving the high dose of BCG by scarification showed relapse in only 1, with follow-up from 30 to 60 weeks, compared to the low-dose BCG group, in which all patients had relapsed and 3 had died by 30 weeks. The latter was comparable to the historical control for this poor-prognosis type of melanoma.

In 1974, our group reported results of BCG immunotherapy in combination with chemotherapy for advanced metastatic malignant melanoma (43). Eighty-nine patients receiving a 5-day course of 5(3,3 dimethyl-l-triazeno)-imidazole-4-carboxamide (DTIC), followed by BCG administered by scarification on days 7, 12, and 17 of each 21-day cycle, were compared to 111 patients treated with DTIC alone in a historical control series. There was an improvement in overall response rate from 14 to 27%, and in patients over the age of 60, the response rate improved from 4 to 25%. When data were analyzed for response by site of involvement, there was no significant improvement in response of liver or lung metastases, while response of regional lymph nodes and/or skin metastases improved from 18 to 55%. There was a significant improvement in the remission duration among the BCG-treated patients, as well as the duration of survival of remission patients and of the overall group.

In 1975, we reported results of a study in which BCG was administered with the objective of prolonging the disease-free interval and survival in patients with stage III malignant melanoma after extirpation of all clinically detectable recurrent regional lymph node disease (72). BCG was administered by scarification weekly for 3 months and then every other week. A low (6×10^7 organisms per scarification) and a high (6×10^8 organisms) dose were compared. The BCG was administered in the deltoid regions and in the thighs by rotation. Several critically important observations were made. First, the high dose of BCG was more effective than the low dose, and indeed, only the high dose significantly improved the disease-free interval and the survival of the overall group. Second, patients with head and neck primaries whose disease had metastasized to cervical lymph nodes did not benefit from postsurgical BCG immunotherapy. In contrast to this, in the recently reported study of Eilber and co-workers (73), a similar program was used. However, BCG for head and neck patients was also placed over the site of the primary lesion, and in this circumstance, patients with head and neck primaries also responded.

In 1976, McKneally and co-workers reported a study of intrapleural BCG immunotherapy in 39 patients with stage I lung cancer (49). Seventeen patients received 10^7 Tice-strain BCG organisms intrapleurally after surgery, followed 2 weeks later by a three-month course of INH. The controls received surgery and INH without BCG. With a median follow-up of approximately one year, there were no recurrences or deaths in the BCG group, while there were 9 recurrences and 5 deaths in the control group of 22 patients. In a much smaller group of patients with stage II and III lung cancer, no significant effect of intrapleural BCG was observed. Finally, in 1976, Yamamura and co-workers (74) reported the effects of immunotherapy of stage I through IV lung cancer with BCG cell-wall skeleton attached to oil droplets (75, 76). Patients received this material intradermally, intrapleurally, intralesionally, and in repeated doses between courses of chemotherapy as the clinical situation indicated. In each stage of lung cancer, the patients receiving this vaccine had improved survival compared to patients not receiving the vaccine. For example, in patients with stage IV disease receiving chemotherapy plus vaccine, 45% were alive at 15 months, compared to a historical control group receiving chemotherapy alone.

The papers reviewed in this section establish many of the principles of immunotherapy with BCG and its subfractions and point the direction for future work in this field. BCG administration can prolong the disease-free interval or remission duration and/or survival in both leukemias and in solid tumors, when a state of minimal residual disease has been established by conventional treatment. Dose of administration is important, and a high dose is more effective. Route of administration is also important, and disease that metastasizes via lymph nodes must be treated along the lymphatic pathways of spread of the disease. Thus BCG must be administered in the region of potential or existing metastatic disease. Presumably, this relates to the need for activation of regional host defense mechanisms.

BCG immunotherapy can be effective even when the tumor burden is large. This contradicts the hypothesis that a minimal tumor burden of under 10^5 cells is required for effective therapy (68). The administration of BCG with chemotherapy can actually improve the overall response rate of metastatic disease. The individual must be immunocompetent to receive the optimal benefit from BCG immunotherapy. Nonviable components of BCG, that is, purified active fractions, may be as or more effective than native BCG. These observations suggest that future work with BCG, as well as with other microbial adjuvants, should emphasize improving the immunocompetence of the patient prior to or concurrent with BCG administration, using the BCG immunotherapy as intensively as possible in terms of dose and schedule, delivering the BCG or BCG derivative to the site of tumor involvement (this suggests intravenous administration in the case of liver, lung, or bone metastases), and finally, the need for preclinical and clinical laboratory work on purified active adjuvant subcomponents of BCG and other microbes.

IMMUNOTHERAPY OF LEUKEMIA WITH BCG

There have now been more than 20 trials of immunotherapy in acute leukemia, and at least 12 of these involved the use of BCG or a BCG derivative (77). Most trials subsequent to the report of Mathé and co-workers have been negative in acute lymphocytic leukemia (ALL), but have been positive in acute myeloid leukemia (AML) and chronic myeloid leukemia (CML). Thus, Heyn and co-workers administered BCG immunotherapy to children with ALL. When in remission, patients received either no further therapy, BCG, or chemotherapy. BCG was inferior to chemotherapy for remission maintenance (78). This study utilized a very low dose of Tice-strain BCG. The British Medical Research Council administered BCG by multipuncture plate subsequent to remission induction with combination chemotherapy for childhood ALL (79). The BCG used was a low dose of Glaxo strain. No effect of immunotherapy was observed. Poplack and co-workers studied BCG immunotherapy added to chemotherapy and craniospinal irradiation for childhood

ALL. Remission maintenance was with chemotherapy or four monthly cycles of chemotherapy alternating with 2 months of immunotherapy with BCG of the Pasteur strain administered by multipuncture plate (80). No effect of BCG immunotherapy was observed. These studies were all in children with ALL. In our own immunotherapy studies with BCG, administered by scarification between courses of chemotherapy, while there was immunotherapeutic benefit in adults with AML, there was no immunotherapeutic benefit in adults with ALL (81). Difficulty in demonstrating efficacy of BCG immunotherapy in ALL may relate to the high response rate and long remission duration induced by chemotherapy alone, and the demonstrated weak immunogenicity of ALL cells.

In contrast to the above, immunotherapy of AML with BCG alone or BCG plus tumor-cell vaccine appears to be effective. Vogler and colleagues randomized 41 patients in remission after chemotherapy to receive BCG of the Tice strain twice weekly for four weeks, followed by chemotherapy, or to receive chemotherapy alone (82). Remission duration was prolonged in the immunotherapy group. In our study (81) we used liquid Pasteur-strain BCG administered approximately weekly by scarification between courses of remission-maintenance chemotherapy, and have demonstrated prolongation of remission and survival. In historical controls, the remission duration was 50 weeks, compared to 91 weeks among patients on chemoimmunotherapy, while the median survival was 60 weeks in the control, compared to greater than 72 weeks among patients on immunotherapy. Recently, we have compared patients receiving remission maintenance, chemoimmunotherapy, plus late chemotherapy consolidation to patients receiving the same, plus BCG immunotherapy. This resulted in a median duration of remission of 160 weeks, compared to 90 weeks on chemotherapy alone (83). It is of great interest that Whittaker and co-workers have compared immunotherapy with intravenous BCG plus chemotherapy to chemotherapy alone in patients with AML (48). Both remission and survival were longer in the group receiving BCG. Second remissions were obtained more easily in patients receiving the BCG.

Nonviable fractions of BCG are apparently also active in immunotherapy of AML. There have now been two reports that remission duration and/or survival can be improved by the administration of MER between courses of remission-maintenance chemotherapy (84, 85).

BCG has also been used in AML in combination with active-specific immunotherapy with allogeneic irradiated leukemic cells. The results of these studies are similar to the use of BCG alone, and the exact role of either component in these combination programs cannot be determined. A series of studies was carried out under the direction of the late Professor Gordon Hamilton-Fairley in London over the last few years (86–88). Remission-induction chemotherapy was given, followed by intermittent combination chemotherapy plus immunotherapy. The latter was irradiated allogeneic leukemic cells plus Glaxo-strain BCG administered by Heaf gun. The dosage of leukemic cells was 10^9 divided between the three extremities not receiving BCG, and the whole program was administered approximately weekly by rotation. It appears that the main effect of this immunotherapy is to increase survival duration from 45 to approximately 78 weeks. In one of these series of studies, it was demonstrated that survival duration was as long with immunotherapy alone as with chemoimmunotherapy. This form of combined immunotherapy, BCG plus tumor-cell vaccine, has also been used in CML (89). Sokal and co-workers administered repeated doses of BCG and a cultured allogeneic leukocyte cell line. Prolongation of disease-free interval and survival was observed (89).

Future developments in BCG immunotherapy of the leukemias should focus on improving immunocompetence of the patient, defining and optimizing the interaction between immunotherapy and chemoimmunotherapy, defining and optimizing the role of other approaches to immunotherapy such as tumor-cell or antigen immunization, investigating the role of purified subfractions of BCG and other microbial agents, and further investigating alternate route of immunotherapy such as intravenous administration.

IMMUNOTHERAPY OF LUNG CANCER WITH BCG

Relatively few studies of immunotherapy of lung cancer with BCG have been carried out. Villasor, in the study mentioned above, included lung cancer patients in his series and observed prolongation of survival when intracutaneous BCG was added to chemotherapy (70). Khadzhiev & Kavaklieva-Dimitrova reported a series of studies in which an aqueous extract of the Russian strain of BCG was administered to patients with inoperable bronchogenic carcinoma, some of whom were receiving other treatment (90). Regression of histologically and radiographically proven tumors occurred in 28% of the patients and was attributed to immunotherapy. The survival rate of patients receiving immunotherapy was superior to that in an untreated control group. The study of McKneally and co-workers, utilizing intrapleural BCG (49), and the study of Yamamura and co-workers, utilizing BCG cell-wall skeleton coated on oil droplets for regional, local, and systemic immunotherapy of lung cancer, have already been outlined (74). In 1974, Edwards & Whitwell reported that a single subcutaneous dose of Glaxo BCG, given 10 days after surgery, improved the survival of patients with various types of lung cancer (91). The two-year survival in squamous cell carcinoma increased from 50 to 62%, and in squamous cell carcinoma with positive nodes, from 33 to 53%. In oat cell carcinoma, the two-year survival increased from 11 to 50% with the addition of BCG. These results were suggestive, but did not reach statistical significance. The authors are now conducting a randomized trial that involves 500 patients, but the result has not yet been reported. Recently, Pines reported a study in which 48 patients with locally advanced squamous cell carcinoma were randomized to receive radiotherapy alone or radiotherapy plus weekly BCG applied with a Heaf gun (92). They used Glaxo BCG at a dosage of 1.2×10^8 organisms per application. None of 23 patients in the control group and 6 of 22 patients in the BCG group were alive at 3 years. Also, distant metastases were significantly less common subsequent to radiotherapy in the BCG group.

Several phase I trials of aerosol administration of BCG have been attempted, but these have been associated with a fairly severe set of pulmonary side effects, and the ultimate utility of this approach to immunotherapy is questionable (93).

In lung cancer, significant advances in immunotherapy have also been made using other modalities, and these should at least be mentioned. Thus, in stage I patients, active specific immunotherapy with tumor antigen mixed with Freund's adjuvant (94) and immunorestorative immunotherapy with levamisole (95) have been shown to be effective. In patients with advanced disease, either lung cancer or with pulmonary metastases of other tumors, intravenous *C. parvum* alone has caused tumor regression (96). In a recent study of locally advanced lung cancer, chemotherapy with adriamycin plus *C. parvum* has proven to prolong survival compared to chemotherapy with adriamycin alone (97). These data suggest that BCG immunotherapy may play an important role in combination with other forms of immunotherapy, and that systemic immunotherapy with BCG or BCG derivatives by the intravenous route should be seriously considered.

IMMUNOTHERAPY OF MALIGNANT MELANOMA WITH BCG

Malignant melanoma has been a subject for immunotherapy for a considerable period of time because of its known immunological associations. Thus, cellular infiltration of the primary tumor and spontaneous regression of cutaneous metastases, as well as spontaneous regression of primary tumors, have been observed, suggesting that this tumor is a particularly responsive one to host control factors (98). Both humoral and cell-mediated immunity to malignant melanoma have been identified in various laboratory studies and high levels of immunity are generally associated with a good prognosis or a state of minimal residual disease (99). Some of the early important immunotherapeutic studies in malignant melanoma that laid the foundations of BCG immunotherapy have already been described. The systemic adjuvant studies of Bluming (71) and the intralesional studies of Morton (69) prompted a series of studies by different investigators of adjuvant immunotherapy with either BCG alone, or BCG plus chemotherapy, usually with DTIC. In these studies, patients with local recurrence or regional lymph node metastases, or even with distant metastatic disease, are rendered free of clinically detectable disease by surgery, after which repeated administration of BCG is carried out, usually by scarification, multipuncture, or Heaf gun technique, either alone or between courses of chemotherapy. Patients with regional lymph node metastases discovered at the time of prophylactic dissection, concurrent with removal of the primary tumor, or patients with clinically recurrent regional lymph node disease months to years after primary surgery, have both been studied.

A variety of investigators, including Gutterman (72), Eilber (73), Morton (99), Ikonopisov (100), Siegler (101), Milton (102), and their colleagues, have reported prolonged disease-free interval and survival. The responsive patients are usually those who are immunocompetent and those receiving the more aggressive regimens of therapy. The usual response is a doubling of the disease-free interval and the survival, with the shapes of the treated and the control curves being the same. For example, in the recent report by Eilber and co-workers (73), the treated group had a 55% three-year disease-free interval, compared to 25% in the control group. In that study, for different sites there was an increase in the percentage of patients free of disease from 30 to 75% for head and neck primaries, 25 to 50% for extremity primaries, and 45 to 70% for trunk primaries. Not all studies have been positive. Pinsky and co-workers have not found an adjuvant effect of BCG, but they used a low dosage and did not administer BCG regional to the primary tumor (103). For patients with metastatic disease, some improvement of response rate to chemotherapy has been observed by the addition of BCG immunotherapy between courses, but the major effect is prolongation of remission duration and survival (43).

BCG immunotherapy has also played an important role when combined with tumor-cell vaccines. BCG mixed with irradiated tumor cells has been used in two studies, one alone, and one in combination with chemotherapy (104, 105). When used alone, a transient, approximately 25%, partial remission rate was observed. When used in combination with DTIC, this regimen led to a 50% remission rate compared to the usual 20 to 25% with DTIC alone. Again in studies of metastatic disease, not all reports have been positive, and some ongoing studies have failed to confirm the additive results of BCG or BCG plus tumor cells in the treatment of metastatic disease.

In contrast, Morton's original observation of the efficacy of intralesional immunotherapy has been extensively confirmed by several investigators including Siegler (101), Minton (106), Pinsky (107), Nathanson (108), and Mastrangelo (109), and their colleagues. BCG subfractions also appear to be active in intralesional immunotherapy. Yamamura and co-workers reported a case in which there was cutaneous metastatic disease, as well as regional lymph node involvement (110). Repeated intralesional injection of BCG cell-wall skeleton attached to oil droplets resulted in regression of measurable disease over several months. In our clinic, Richman and co-workers (111) have preliminary observations suggesting that patients with advanced metastatic disease show approximately a 20% response rate of locally injected nodules, either with BCG cell-wall skeleton on oil droplets alone, or when combined with the P3 (trehalose dimycolate) fraction of BCG. Because of these observations, and observations made in the guinea pig hepatoma model, described above, intralesional immunotherapy of primary melanoma as a method to improve the ultimate cure rate from surgery is very attractive. This is currently under study with BCG. However, Everall and co-workers recently carried out intralesional immunotherapy with vaccinia virus before primary surgery for levels IV and V primary melanomas (112). There was a prolongation in the disease-free interval in the immunotherapy group, with 50% of the controls, but only 5% of the treated patients relapsing in 24 months. At 54 months, however, there was no significant difference between the groups. Thus, results of immunotherapy with melanoma are encouraging. However, a doubling of the disease-free interval and survival and a prolongation of remission rate are not adequate in a disease where virtually 100% of patients with 5 or more positive nodes will ultimately expire, and where the current chemotherapy response rate hovers around 20%. Research in immunotherapy of melanoma with BCG is currently focused on BCG subcomponents, intralesional immunotherapy of the primary tumor before surgery, more intensive use of BCG, and attempts to use either BCG or some other bacterial adjuvant to reach systemic areas of involvement, either gross or microscopic. In this regard, daily intravenous or continuous intravenous immunotherapy with *C. parvum* is showing promise (96).

IMMUNOTHERAPY OF OTHER SOLID TUMORS WITH BCG

A variety of other solid tumors have been treated with either BCG or BCG derivatives. One of the most interesting areas is carcinoma of the colon. Mavligit and co-workers have done a study in 76 patients with Dukes's class C colon cancer, in which the patients were randomized to receive either BCG by scarification weekly for 3 months and every other week thereafter, or to receive a 5-day course of 5-fluorouracil (5-FU) every three weeks, with weekly BCG between courses (113). These patients were compared to historical controls. There was a significant increase in the disease-free interval and survival of these patients. The regimen was tolerated with virtually no systemic side effects. Both regimens were equivalent in their efficacy. The greatest number of relapses were in patients with rectosigmoid carcinoma who relapsed locally. Thus, even though BCG was given into the lower extremities, it apparently did not affect microscopic disease in the pelvic area. The remaining relapses were mainly in the liver. We have now devised a new regimen in which the BCG immunotherapy will be supplemented by an early course of daily intravenous *C. parvum,* with the problem of local relapses to be attacked by local radiotherapy. Also, the doses of 5-FU will be increased into the moderately myelosuppressive range. While no data are available, this multimodality approach offers promise.

In a phase I study of immunotherapy with MER, Moertel and co-workers observed that 3 out of 38 patients with metastases of colon carcinoma to lung or liver showed transient but definite measurable partial remission (114).

In metastatic breast carcinoma, major advances have recently been made in chemotherapy. One effective regimen is a combination of 5-FU, adriamycin, and cyclophosphamide (referred to as FAC). This induces a 75% complete and partial remission rate and a 25% complete remission rate (115). Administration of 6×10^8 BCG organisms on days 9, 13, and 17 between 21-day cycles of chemotherapy has significantly prolonged the duration of complete plus partial remission, and has significantly prolonged survival (116). This regimen is also being used in the treatment of patients with stage II breast cancer after surgery. While preliminary, it appears that this is a highly effective regimen in the prevention of recurrent disease.

BCG immunotherapy has also been used in head and neck cancer. Donaldson reported that a combination of methotrexate, isoniazid, and BCG increased the complete and partial remission rate in patients with recurrent or advanced squamous cell carcinoma from the usual 35% to approximately 80% (117). In our clinic, Richman and co-workers have observed that for patients with distant metastases of head and neck cancer, combination chemotherapy, plus BCG given by scarification approximately weekly, had no effect on the response rate or response duration, but significantly prolonged survival (118).

In another study gastrointestinal carcinoma, in patients with advanced but partially operable carcinomas of various gastrointestinal sites, intraperitoneal BCG, followed by boosting with oral BCG, caused apparent regression of tumor in some patients (50). This study is hard to evaluate because of the multiplicity of histological types treated, their variability of clinical extent of disease, the difficulty in evaluating postoperative intraabdominal disease, and the lack of a control group. However, this study has demonstrated the feasibility of intraperitoneal BCG administration, and the rationale for local-regional immunotherapy with BCG was discussed above.

Several studies have been conducted involving BCG immunotherapy in malignant lymphoma. Sokal and co-workers reported a study in which 50 patients with IA through 2A malignant lymphoma of various histological types previously treated with radiotherapy were given a single intradermal injection of 0.1 ml of Tice-strain BCG or were not treated (119). The relapse rate at 6–8 years was 77% in the controls and 61% in the treated; the mean time to the first new lesion was 10.6 months in the control, and 25.9 months in the treated. This study is difficult to interpret because of marked variability between the two groups in various prognostic factors and the fact that staging procedures were inadequate at the time of the study. Recently, Mathé and co-workers treated a group of patients with leukemic lymphosarcoma with chemotherapy. Of the 18 who entered remission, 9 were treated with active immunotherapy, including BCG, *C. parvum*, *C. granulosum*, and irradiated allogeneic leukemic cells (120). Five of the 9 remained free of disease for 13–66 months, while 7 of the 9 in the control group have already died.

In soft-tissue sarcoma, Townsend and co-workers treated patients with BCG immunotherapy and a cultured sarcoma tumor-cell vaccine after surgical extirpation of their disease (121). Sixty-one percent of 18 patients with localized soft-tissue sarcoma who received immunotherapy were free of disease at the time of the report, compared to only 33% of the 15 treated with surgery alone. Of the patients who relapsed, the disease-free interval was 15 months in the treated patients and 7.3 months in the control. This result is quite analogous to that seen in other disease categories such as leukemia, lymphoma, melanoma, and carcinoma of the colon.

A most interesting preliminary study in bladder carcinoma has been reported by Eidinger & Morales (122). These investigators used intravesicle BCG immunotherapy plus BCG administered intracutaneously in the thighs in patients with recurrent bladder cancer who had been previously treated with fulguration alone or fulguration plus the installation of thiotepa and epodyl. The recurrence rate prior to immunotherapy averaged 6 months, with 5 new lesions appearing during that period of time. After the BCG treatment, patients had an average of 0 new lesions in 5 months. Overall, there were 35 lesions in 67 months of follow-up in 7 patient prior to treatment, and none in 47 months of follow-up after treatment. This strongly suggests the potential efficacy of regional immunotherapy with BCG. There are other examples of BCG immunotherapy, most of them too preliminary for definitive evaluation, but nonetheless encouraging. This overview of some of the more encouraging studies should give the reader a reasonable concept of the current status of the use of this agent in cancer treatment.

SUBCOMPONENTS OF BCG FOR CANCER IMMUNOTHERAPY

There is currently a great deal of interest in purified subcomponents and nonviable fractions of BCG for immunotherapy. Interest in these stems from two basic motivations. First, there are serious complications and side-effects of immunotherapy with viable BCG. Possibly purified subcomponents could avoid these. Second, purifying active subcomponents should result in a more active preparation and in a preparation that can be administered through a wider variety of routes more safely. The subcomponent preparations currently being used clinically are the methanol-extraction residue of BCG (MER), and various defined products, including purified BCG cell-walls, purified BCG cell-wall skeleton (peptidogalactan), and purified cell-wall lipid produced by specialized column fractionation, referred to as P3 (trehelose dimycholate).

MER is made by methanol extraction of phenol-killed, acid-washed BCG organisms and represents the aggregate or particulate portion of partially delipidated cell-wall skeletons of the BCG organism (123). MER was initially developed as a potential adjuvant to increase resistance to mycobacterial and other infection. Subsequently, it was discovered that it had potent antitumor activity (123). The material increases resistance to viral leukemogenesis and increases resistance to transplantation of carcinogen-induced tumors. It can prolong survival following chemotherapy or radiotherapy of experimental tumors, although it usually does not produce cures. Of interest, administration at sites distal to the tumor is as good as intralesional administration. Its mechanism of action is similar to that of BCG. MER has been shown to prolong remission duration and survival in acute leukemia in two studies (84, 85), and to induce transient remissions in patients with metastatic colon cancer (114). Our group has recently conducted a study comparing the efficacy of BCG and MER in the chemoimmunotherapy of metastatic breast cancer. Equivalent activity was demonstrated and produced superior survival compared to chemotherapy alone (124). The other BCG fractions have mainly been studied by Ribi & Lederer. BCG cell-wall is made by lysis of BCG, and cell-wall skeleton is made by proteolysis and extraction with organic solvents (125). The material extracted with organic solvents results in a series of glycolipids, including wax-D and cord factor, a purified subcomponent of which is P3. The best source of cell-wall skeleton is *M. phlei,* while the best source of P3 is the aoyma B strain of *Mycobacterium tuberculosis*. BCG cell-wall skeleton has been used extensively in only one study of immunotherapy of lung cancer outlined above (74), and has been (also outlined above) used in a number of studies of intralesional immunotherapy in melanoma (110, 111). Both of these agents offer great promise for the future of immunotherapy. Indeed, in recent animal-model trials these materials, as well as more highly refined microbial products such as detoxified mutant endotoxin plus P3 on oil droplets, are proving highly effective (128).

UNTOWARD REACTIONS AND COMPLICATIONS

The usual or normal reactions to BCG administration were described above. However, it would be worthwhile to outline the unusual but significant, serious complications of this form of treatment encountered. We might first consider local complications. When BCG is administered intradermally or subcutaneously, the local induced lesion heals after a month or two. However, in a small fraction of patients, perhaps 5%, a chronic local draining lesion develops, which ulcerates and may persist for many months, or even years. Such a lesion results in scarring. When BCG is administered by scarification or multipuncture plate, very severe local reactions occur in less than 5%. These are associated with pain, blistering, oozing, or even bleeding; these lesions, rather than healing in two weeks, may take several months to heal. Although very rare, secondary bacterial infection of scarified or multipuncture-treated areas may occur. These require local treatment with soaks and peroxide, and the surrounding cellulitis may require systemic antibiotic treatment.

Several types of systemic reactions can also occur. While it has not been described for either scarification, intradermal, or subcutaneous administration, the intralesional administration of BCG has resulted in at least two cases of an anaphylactoid syndrome, followed by death (44). Approximately 2% of patients receiving repeated BCG administration in our clinic develop generalized cutaneous allergic reactions that may persist for weeks, which are associated with systemic symptoms and may require corticosteroids and antituberculous agents in combination for their control. We have observed erythema multiforme, generalized maculopapular eruption, erythema nodosum, exfoliative dermatitis, psoriatiform eruption, and papulonecrotic tuberculid as the types of systemic cutaneous reactions. The management of these includes cessation of BCG immunotherapy and observation. If the patient progresses over several days or develops significant systemic symptoms, antituberculous therapy is started, and the patient is given a short course of corticosteroids with rapidly tapering doses over a period of five days. Antituberculous therapy is continued for at least one month and the patient is considered to be unacceptable for BCG immunotherapy in the future. Such patients are usually placed on subcutaneous *C. parvum* as an alternative.

Finally, patients receiving BCG immunotherapy may develop systemic BCG disease. This is manifested by fever, malaise, night sweats, weight loss, and is associated with rather severe changes in hepatic function. Diagnosis is made by biopsy and culture of bone marrow and/or liver. In approximately 3000 patients who have received BCG by scarification in our clinic, only 2 have developed systemic BCG disease, even though BCG has been administered to several hundred patients who were severly myelo- and immunosuppressed during chemotherapy for acute leukemia. In contrast, the incidence of systemic BCG disease may be as high as 50%, although it is usually approximately 5% among patients receiving relatively large doses of BCG intralesionally. Treatment of systemic BCG disease involves the usual combination chemotherapy for tuberculosis. Response can be anticipated in every case since the organism is sensitive.

Many investigators have expressed concern over the possibility of tumor enhancement induced by BCG, when used either alone or in combination with chemotherapy or other forms of immunotherapy. Since this is a real possibility and has been demonstrated in animal models and in our own experience, we suspect that tumor enhancement occurred in one small group of melanoma patients seen in our clinic who received one low-viability, low-dose lot of BCG from a particular supplier. Carefully controlled studies are required to rule out this possibility.

LOCAL-REGIONAL IMMUNOTHERAPY WITH BCG

Local-regional immunotherapy with BCG and its derivatives has great potential. Animal model studies already discussed have indicated that primary intralesional injection of tumor can cause regression of that injected nodule, inhibition of the development of regional lymph node metastases, and increased generalized or systemic tumor immunity (60). The requirements for this effect are either intimate contact between the BCG and the tumor cells, or that the tumor and the BCG be in the same regional lymph node drainage. The adjuvant effect and the induction of specific tumor immunity therefore probably involve interaction between tumor antigen and products of the BCG organisms at the lymph node level. At the local tumor site after BCG inoculation, there is an acute followed by a chronic inflammatory response, the latter stages of which probably involve the development of delayed-type hypersensitivity (58). Killing of tumor cells is the result of a bystander effect of this delayed hypersensitivity reaction, as well as specific tumor immunity.

Direct intralesional immunotherapy has been applied to metastatic nodules of breast cancer and melanoma (103), as described above, and probably would be effective against most tumors. Usually, the injected nodule regresses while distant metastatic nodules do not regress. Regression is usually not observed in severely immunologically incompetent hosts. This suggests that efforts to restore immunocompetence might precede intralesional therapy with BCG. BCG subcomponents are also active in intralesional therapy. This includes PPD treatment of squamous tumors, and BCG cell-wall skeleton and MER treatment of melanoma nodules (110).

Perhaps potentially more important in the long run is not treatment of metastatic nodules, but local regional therapy of early or primary cancer. This is likely to be effective in those cases in which the tumor first metastasizes by regional lymph nodes, and in which microscopic metastases are present at the time of surgery, making the prognosis poor. Induction of vigorous tumor immunity by local regional immunotherapy under this circumstance should markedly improve the patient's prognosis. This has already been demonstrated in melanoma (112) and in lung cancer (49), and very much needs to be widely applied in a variety of other solid tumors. Local-regional immunotherapy for tumors that metastasize to liver or to lung via hematogenous routes could be treated by local-regional immunotherapy by means of intravenous or intraarterial administration of various microbial fractions. One experiment is already underway in which intralesional BCG is being given for primary levels IV or V malignant melanoma, in hopes of improving the prognosis (96).

SUMMARY AND CONCLUSIONS

Immunotherapy with BCG is a clinical discipline that is approximately 10 years old. The cutaneous application of BCG has shown some efficacy in the treatment of leukemia; lymphoma; melanoma; sarcoma; breast, colon, lung, bladder, head and neck cancer; and other tumors. The usual activity of BCG is prolongation of remission duration and survival in association with conventional therapy. BCG has been shown to be active in advanced as well as minimal residual disease. Improvement in response rates occurs occasionally, but has not been generally observed. The effect of BCG is usually modest, and it certainly is not a panacea. Indeed, not all studies in a given disease category have yielded positive immunotherapeutic results, although in some of these inadequate dose, route, or schedule of administration of BCG may explain the negative result.

While BCG immunotherapy appears to be indicated in a variety of disease categories, since it usually does not improve the response rate and has modest therapeutic efficacy, it may well be superseded by more effective approaches in the future. These would include other approaches to immunotherapy, particularly systemic immunotherapy, as well as more effective approaches to adjuvant treatment with chemotherapy or immunotherapy in that setting. Future prospects for immunotherapy with BCG include refinements in dose, route, and schedule of administration. We can also anticipate improvement in product characteristics, including standardization, viability, and lack of contaminating undesirable components. Most important will be the development of purified subfractions and the active molecular components of the whole organism. These would include components with adjuvant activity, cross-reactive antigenicity and immunogenicity with the tumor, and perhaps direct antitumor activity. Also, either with intact BCG organisms, or more likely with a purified component of BCG as well as other bacterial adjuvants, systemic immunotherapy will be explored actively. This will require intravenous administration of the materials. Finally, the potential of early or primary local-regional immunotherapy has barely begun to be explored. Preliminary results from both animal models and clinical trials suggest that this approach with BCG, other microbial adjuvants, or one of the subfractions, may alter the response rates to surgery or radiotherapy and the ultimate cure rates of conventional therapy.

Literature Cited

1. Baldwin, R. W., Price, M. R. 1975. Neoantigen expression in chemical carcinogenesis. In *Cancer, A Comprehensive Treatise,* ed. F. F. Becker, 4:353. New York:Plenum
2. Herberman, R. B., Hollinshead, A., Char, D., Oldham, R., McCoy, J., Cohen, M. 1975. In vivo and in vitro studies of cell-mediated immune response to antigens associated with malignant melanoma. *Behring Inst. Mitt.* 56:131
3. Golub, S. H. 1975. Host immune response to human tumor antigens. See Ref. 1, p. 259
4. Carswell, E. A., Old, L. J., Kassel, R. L., Green, S., Fiore, N., Williamson, B. 1975. An endotoxin-induced serum factor that causes necrosis of tumors. *Proc. Natl. Acad. Sci.* 72:3666–70
5. Ruddle, N., Waksman, D. H. 1967. Cytotoxic effect of lymphocyte-antigen interaction in delayed hypersensitivity. *Science* 157:1060–62
6. Kaplan, A. M., Morahan, P. S. 1976. Macrophage mediated tumor cell cytotoxicity. *Ann. NY Acad. Sci.* 276:134–45
7. Hanna, M. G., Bucana, C., Hobbs, B. 1976. Morphologic aspects of tumor cell cytotoxicity by effector cells of the macrophage histiocyte compartment. In *The Macrophage in Neoplasia,* ed. M. Fink. New York:Academic. In press
8. Morton, D. L., Eilber, F. R., Joseph, W. L. 1970. Immunological factors in hu-

man sarcomas and melanomas: A rational basis for immunotherapy. *Ann. Surg.* 172:740–49

9. Hamlin, I. M. E. 1968. Possible host resistance in carcinoma of the breast: A histological study. *Br. J. Cancer* 22: 383–401
10. Alexander, P., Eccles, S. A., Gauci, C. L. L. 1976. The significance of macrophages in human and experimental tumors. *Ann. NY Acad. Sci.* 276:124–33
11. Hersh, E. M., Mavligit, G. M., Gutterman, J. U. 1976. Immunotherapy of leukemia. *Med. Clin. North Am.* 60: 1019–42
12. Hersh, E. M., Gutterman, J. U., Mavligit, G. M. 1975. Cancer and host defense mechanisms. In *Pathobiology Annual,* ed. H. L. Ioachim, p. 133. New York:Appleton-Century-Crofts
13. Baldwin, R. W., Price, M. R., Robins, R. A. 1973. Inhibition of hepatoma-immune lymph node cell cytotoxicity by tumour-bearer serum and solubilized hepatoma antigen. *Int. J. Cancer* 11:527–35
14. Rossen, R. D., Reisberg, M., Hersh, E. M., Gutterman, J. U. 1976. The C19 binding test for soluble immune complexes: clinical correlation obtained in patients with cancer and other diseases. *J. Natl. Cancer Inst.* In press
15. Tripodi, D., Parks, L. C., Brugmans, J. 1973. Drug-induced restoration of cutaneous delayed hypersensitivity in anergic patients with cancer. *N. Engl. J. Med.* 289:354–57
16. *Thymus Factors in Immunity.* 1975. In *New York Academy of Sciences,* ed. H. Friedman, 249:1–547
17. Hitzig, W. H., Grob, P. J. 1974. Therapeutic use of transfer factor. In *Progress in Clinical Immunology,* ed. R. F. Schwartz, p. 69. New York:Grune & Stratton
18. Pilch, Y. H., Ramming, K. P., Deckers, P. J. 1971. Transfer of tumor immunity with RNA. *Isr. J. Med. Sci.* 7:246–58
19. Wright, P. W., Hellström, K. E., Hellström, I. E., Bernstein, I. D. 1976. Serotherapy of malignant disease. *Med. Clin. North Am.* 60:607–22
20. Hersh, E. M., Gutterman, J. U., Mavligit, G. M. 1973. Immunotherapy of cancer. In *Major Problems of Childhood Cancer.* Springfield, Ill:Thomas
21. Fefer, A., Einstein, A. B., Cheever, M. A. 1976. Models for syngeneic adoptive chemoimmunotherapy of murine leukemias. *Ann. NY Acad. Sci.* 276:573–83
22. Fefer, A. 1969. Immunotherapy and chemotherapy of Moloney sarcoma virus-induced tumors in mice. *Cancer Res.* 29:2177–83
23. Zbar, B., Bernstein, I. D., Bartlett, G. L. 1972. Immunotherapy of cancer: regression of intradermal tumors and prevention of growth of lymph node metastases after intralesional injection of living *Mycobacterium bovis. J. Natl. Cancer Inst.* 49:119–30
24. Oettgen, H. F., Pinsky, C. M., Delmonte, L. 1976. Treatment of cancer with immunomodulators: Corynebacterium parvum and levamisole. *Med. Clin. North Am.* 60:511–37
25. Laucius, J. F., Bodurtha, A. J., Mastrangelo, M. J. 1974. Bacillus Calmette-Guerin in the treatment of neoplastic disease. *J. Reticuloendothel. Soc.* 16: 347–73
26. Mackaness, G. B., Auclair, D. J., Lagrange, P. H. 1973. Immunopotentiation with BCG. I. Immune response to different strains and preparations. *J. Natl. Cancer Inst.* 51:1655–67
27. Mackaness, G. B., personal communication
28. Palmer, C. E., Shaw, L. W., Comstock, G. W. 1958. Vaccination: Community trials. *Am. Rev. Respir. Dis.* 77:877–907
29. Sutherland, I. 1972. State of the art in immunoprophylaxis of tuberculosis. In *State of Immunization in Tuberculosis* in 1971, USDEW #(USDHEW #NIH 72-68), p. 113
30. Halpern, B. N., Biozzi, C., Stiffel, C., Mouton, D. 1958. Correlation between phagocytic activity of the reticuloendothelial system and the production of antibacterial antibodies. *CR Soc. Biol.* 152:758–61
31. Howard, J. G., Biozzi, G., Halpern, B. N., Stiffel, C., Mouton, D. 1959. The effect of *Mycobacterium tuberculosis* (BCG) infection on the resistance of mice to bacterial endotoxin and Salmonella enteritidis infection. *Br. J. Exp. Pathol.* 40:281–90
32. Stjernsward, J. 1966. Effect of BCG and/or methocholanthrene on the antibody-forming cells measured at the cellular level by a hemolytic plaque test. *Cancer Res.* 26:1591–94
33. Miller, T. E., Mackaness, G. B., Lagrange, P. H. 1973. Immunopotentiation with BCG: II. Modulation of the response to sheep red blood cells. *J. Natl. Cancer Inst.* 51:1669–76
34. Hibbs, J. B. 1974. Discrimination between neoplastic and non-neoplastic

cells in vitro by activated macrophages. *J. Natl. Cancer Inst.* 53:1487–92

35. Campbell, P. 1976. Immunocompetent cells in resistance to bacterial infections. *Bacteriol. Rev.* 40:284–313
36. Bucana, C., Hanna, M. G. Jr. 1974. Immunoelectronmicroscopic analysis of surface antigens common to Mycobacterium bovis (BCG) and tumor cells. *J. Natl. Cancer Inst.* 53:1313
37. Hanna, M. G. Jr., Peters, L., Gutterman, J. U., Hersh, E. M. 1976. Evaluation of BCG administered by scarification for immunotherapy of metastatic hepatocarcinoma in the Guinea pig. *J. Natl. Cancer Inst.* 56:1013–17
38. Hersh, E. M., Gutterman, J. U., Mavligit, G. M., Mountain, C. W., McBride, C. M., Burgess, M. A. 1976. Immunocompetence, immunodeficiency, and prognosis in cancer. *Ann. NY Acad. Sci.* 276:386–406
39. Mavligit, G. M., Gutterman, J. U., Hersh, E. M. In press
40. Hersh, E. M., Gutterman, J. U., Mavligit, G. M., unpublished observations
41. Farquhar, D., Loo, T. L., Gutterman, J. U., Hersh, E. M., Luna, M. A. 1976. Inhibition of drug-metabolizing enzymes in the rat after Bacillus Calmette-Guerin treatment. *Biochem. Pharmacol.* 25:1529–35
42. Mackaness, G. B., personal communication
43. Gutterman, J. U., Mavligit, G. M., Gottlieb, J. A., Burgess, M. A., McBride, C. E., Einhorn, L., Freireich, E. J., Hersh, E. M. 1974. Chemoimmunotherapy of disseminated malignant melanoma with dimethyl triazeno imidazole carboxamide and Bacillus Calmette-Guerin. *N. Engl. J. Med.* 291:592–97
44. Sparks, F. C. 1976. Hazards and complications of BCG immunotherapy. *Med. Clin. North Am.* 60:499–509
45. Richman, S. P., Mavligit, G. M., Wolk, R., Gutterman, J. U., Hersh, E. M. 1975. Epilesional scarification: Preliminary report of a new approach to local immunotherapy with BCG. *J. Am. Med. Assoc.* 234:1233–35
46. Falk, R. E., Mann, P., Langer, B. 1973. Cell-mediated immunity to human tumors. Abrogation by serum factors and non-specific effects of oral BCG therapy. *Arch. Surg.* 107:261–65
47. Baldwin, R. W., Hopper, D. G., Pimm, M. V. 1975. Influence of orally administered BCG on growth of transplanted rat tumours. *Br. J. Cancer* 31:124–28
48. Whittaker, J. A., Slater, A. J. 1977. The immunotherapy of acute myelogenous leukemia using intravenous BCG. In *Immunotherapy of Cancer: Present Status of Trials in Man*, ed. W. Terry, D. Windhurst. In press
49. McKneally, M. F., Maver, C., Kausel, H. W. 1976. Regional immunotherapy of lung cancer with intrapleural B.C.G. *Lancet* 1:377–85
50. Falk, R. E., MacGregor, A. B., Landi, S., Ambus, U., Langer, B. 1976. Immunostimulation with intraperitoneally administered Bacille Calmette Guerin for advanced malignant tumors of the gastrointestinal tract. *Surg. Gynecol. Obstet.* 142:363–68
51. Halpern, B. N., Biozzi, G., Stiffel, C., Mouton, D. 1959. Effect of stimulation of the reticuloendothelial system by inoculation of the Calmette-Guerin bacillus on the development of the atypical Guerin T-S epithelioma in the rat. *CR Soc. Biol.* 153:919–23
52. Old, L. J., Clark, D. A., Benacerraf, B. 1959. Effect of Bacillus Calmette-Guerin infection on transplanted tumours in the mouse. *Nature* 184:291–92
53. Mathé, G., Halle-Pannenko, O., Bourut, C. 1973. BCG in cancer immunotherapy: Results obtained with various BCG preparations in a screening study for systemic adjuvants applicable to cancer immunoprophylaxis or immunotherapy. *Natl. Cancer Inst. Monogr.* 39:107–12
54. Zbar, B., Bernstein, I. D., Rapp, H. J. 1971. Suppression of tumor growth at the site of infection with living Bacillus Calmette-Guerin. *J. Natl. Cancer Inst.* 46:831–39
55. Baldwin, R. W., Pimm, M. V. 1973. BCG immunotherapy of plumonary growths from intravenously transferred rat tumour cells. *Br. J. Cancer* 27:48
56. Amiel, J. L., Berardet, M. 1974. Factor time for active immunotherapy after cytoreductive chemotherapy. *Eur. J. Cancer* 10:89–91
57. Piessens, W. F., Lachapelle, F. L., Legros, N. 1970. Facilitation of rat mammary tumour growth by BCG. *Nature* 228:1210–11
58. Hanna, M. G. Jr., Zbar, B., Rapp, H. J. 1972. Histopathology of tumor regression after intralesional injection of *Mycobacterium bovis*. I. Tumor growth and metastases. *J. Natl. Cancer Inst.* 48:1441–55
59. Fidler, I. J., Buomen, M. B., Hanna, M. G. 1977. Characterization of in vitro

reactivity of BCG-treated guinea pigs to syngeneic line 10 hepato carcinoma. *Cancer Immunol. Immunother.* In press

60. Rapp, H. J. 1973. A Guinea pig model for tumor immunology. A summary. *Isr. J. Med. Sci.* 9:366–74
61. Hanna, M. G. Jr., Peters, L., Fidler, I. J. In press
62. Davignon, L., Lemonde, P., Robillard, P. 1970. BCG vaccination and leukemia mortality. *Lancet* 2:638
63. Rosenthal, S. R., Crispen, R. G., Thorne, M. G. 1972. BCG vaccination and leukemia mortality. *J. Am. Med. Assoc.* 222:1543–44
64. Hoover, R. N. 1976. Bacillus Calmette-Guerin vaccination and cancer prevention: A critical review of the human experience. *Cancer Res.* 36:652–54
65. Kinlen, L. J., Pike, M. C. 1971. BCG vaccination and leukaemia. Evidence of vital statistics. *Lancet* 2:398–402
66. Waaler, H. T. 1971. BCG and leukaemia mortality. *Lancet* 2:1314
67. Mathé, G., Amiel, J. L., Schwarzenberg, L., Schneider, M., Cattan, A., Schlumberger, J. R., Hayat, M., De Vassal, F. 1969. Active immunotherapy for acute lymphoblastic leukaemia. *Lancet* 1:697–99
68. Mathé, G., Pouillart, P., Lapeyranque, F. 1969. Active immunotherapy of L-1210 leukaemia applied after the graft of tumour cells. *Br. J. Cancer* 23:814–24
69. Morton, D. L., Eilber, F. R., Malmgren, R. A., Wood, W. C. 1970. Immunological factors which influence response to immunotherapy in malignant melanoma. *Surgery* 68:158–64
70. Villasor, R. 1965. The clinical use of BCG vaccination in stimulating host resistance to cancer: Phase II. Immunochemotherapy in advanced cancer. *J. Philipp. Med. Assoc.* 41:619–32
71. Bluming, A. Z., Vogel, C. L., Ziegler, J. L. 1972. Immunological effects of BCG in malignant melanoma: Two modes of administration compared. *Ann. Intern. Med.* 76:405–11
72. Gutterman, J. U., Mavligit, G. M., Burgess, M. A., Cardenas, J. O., Blumenschein, G. R., Gottlieb, J. A., McBride, C. M., McCredie, K. B., Bodey, G. P. Sr., Rodriquez, V., Freireich, E. J., Hersh, E. M. 1976. Immunotherapy of human solid tumors and acute leukemia with BCG: Prolongation of disease-free interval and survival. *Cancer Immunol. Immunother.* 1:99–107
73. Eilber, F. R., Morton, D. L., Holmes, C., Sparks, F. C., Ramming, K. P. 1976. Adjuvant immunotherapy with BCG in treatment of regional-lymph-node metastases from malignant melanoma. *N. Engl. J. Med.* 294:237–40
74. Yamamura, Y., Azuma, I., Taniyama, T., Sugimura, K., Hirao, F., Tokuzen, R., Okabe, M., Nakahara, W., Yasumoto, K., Ohta, M. 1976. Immunotherapy of cancer with cell-wall skeleton of *Mycobacterium bovis* BCG: experimental and clinical results. *Ann. NY Acad. Sci.* 277:209–27
75. Ribi, E., Meyer, T. J., Azuma, I. 1973. Mycobacterial cell wall components in tumor suppression and regression. *Natl. Cancer Inst. Monogr.* 39:115–19
76. Zbar, B., Ribi, E., Meyer, T. 1974. Immunotherapy of cancer:Regression of established intradermal tumors after intralesional injection of mycobacterial cell walls attached to oil droplets. *J. Natl. Cancer Inst.* 52:1571–77
77. Hersh, E. M., Gutterman, J. U., Mavligit, G. M. 1976. Immunotherapy of leukemia. *Med. Clin. North Am.* 60: 1019–42
78. Heyn, R. M., Joo, P., Karon, M., Nesbit, M., Shore, N., Breslow, N., Weiner, J., Reed, A., Hammond, D. 1975. BCG in the treatment of acute lymphocytic leukemia. *Blood* 46:431–42
79. Medical Research Council. 1971. Preliminary report to the Medical Research Council by the Leukemia Committee and the Working Party on Leukemia in Childhood: Treatment of acute lymphoblastic leukemia. *Br. Med. J.* 4:189
80. Poplack, D. G., Graw, R. G., Pomeroy, T. C., Henderson, E. S., Leventhal, B. G. 1975. Chemotherapy versus chemotherapy and immunotherapy in childhood acute lymphocytic leukemia. *Proc. Am. Assoc. Cancer Res.* 16:230
81. Gutterman, J. U., Rodriguez, V., Mavligit, G. M., Burgess, M. A., Gehan, E., Hersh, E. M., McCredie, K. B., Reed, R. C., Smith, T., Bodey, G. P. Sr., Freireich, E. J. 1974. Chemoimmunotherapy of adult acute leukaemia prolongation of remission in myeloblastic leukemia with BCG. *Lancet* 2:1405–9
82. Vogler, W. R., Chan, Y. K. 1974. Prolonging remission in myeloblastic leukaemia by Tice-strain Bacillus Calmette-Guerin. *Lancet* 2:128–31
83. Gutterman, J. U., Mavligit, G. M., Burgess, M. A., Cardenas, J. O., Blumenschein, G. R., Gottlieb, J. A., McBride,

C. M., McCredie, K. B., Bodey, G. P. Sr., Rodriguez, V., Freireich, E. J., Hersh, E. M. 1976. Immunotherapy of human solid tumors and acute leukemia with BCG: Prolongation of disease-free interval and survival. *Cancer Immunol. Immunother.* 1:99–107

84. Weiss, D. W., Stupp, Y., Many, N., Izak, G. 1975. Treatment of acute myelogenous leukemia (AML) patients with the MER tubercle Bacillus fraction. A preliminary report. *Transplant Proc.* 7:545
85. Cuttner, J., Holland, J. F., Bekesi, J. G., Ramachander, K., Donovan, P. 1975. Chemoimmunotherapy of acute myelogenous leukemia. *Proc. Am. Assoc. Cancer Res.* 16:264
86. Crowther, D., Powles, R. L., Bateman, C. J. T., Beard, M. E. J., Gauci, C. L., Wrigley, P. F. M., Malpas, J. S., Hamilton-Fairley, G., Scott, R. B. 1973. Management of adult acute myelogenous leukaemia. *Br. Med. J.* 42:131–37
87. Hamilton-Fairley, G. 1975. Immunotherapy in the management of leukaemia. *Br. J. Haematol.* 31:181–92
88. Powles, R. L. 1976. Immunotherapy in the management of acute leukemia. *Br. J. Haematol.* 32:145–49
89. Sokal, J. E., Aungst, C. W., Grace, J. T. 1973. Immunotherapy in well-controlled chronic myelocytic leukemia. *NY State J. Med.* 73:1180–85
90. Khadzhiev, S., Kavaklieva-Dimitrova, Y. 1971. Treatment of bronchial cancer patients with a water-saline extract of BCG. *Vopr. Onkol.* 17:15–57
91. Edwards, F. R., Whitwell, F. 1974. Use of BCG as an immunostimulant in the surgical treatment of carcinoma of the lung. *Thorax* 29:654–58
92. Pines, A. 1976. A 5-year controlled study of BCG and radiotherapy for inoperable lung cancer. *Lancet* 1:380–81
93. Garner, F. B., Meyer, C. A., White, D. S., Lipton, A. 1975. Aerosol BCG treatment of carcinoma metastatic to the lung: A phase I study. *Cancer* 35: 1088–94
94. Stewart, T. H. M., Hollinshead, A. C., Harris, J. E., Belanger, R., Crepeau, A., Hooper, G. D., Sachs, H. J., Klaasser, D. J., Hirte, W., Rapp, E., Crook, A. F., Orizaga, M., Sengar, D. P. S., Ramon, S. 1976. Immunochemotherapy of lung cancer. *Ann. NY Acad. Sci.* 277:436–66
95. Study Group for Bronchogenic Carcinoma. 1975. Immunopotentiation with levamisole in resectable bronchogenic carcinoma: A double-blind controlled trial. *Br. Med. J.* 3:461–64
96. Israel, L., Edelstein, R., Depierre, A., Dimitrov, N. 1975. Brief communication: Daily intravenous infusions of Corynebacterium parvum in twenty patients with disseminated cancer: A preliminary report of clinical and biologic findings. *J. Natl. Cancer Inst.* 55:29–33
97. Dimitrov, N., personal communication
98. Morton, D. L., Eilber, F. R., Holmes, E. C., Sparks, F. C., Ramming, K. 1976. BCG immunotherapy as a systemic adjunct to surgery in malignant melanoma. *Med. Clin. North Am.* 60:431–39
99. Morton, D. L., Holmes, E. C., Eilber, F. R., Wood, W. C. 1971. Immunological aspects of neoplasia: A rational basis for immunotherapy. *Ann. Intern. Med.* 74:587–604
100. Ikonopisov, R. L. 1975. The use of BCG in the combined treatment of malignant melanoma. In *Int. Symp. Immunol. React. Melanoma Antigens,* ed. K. Heide, p. 206. Frankfurt:Behringwerke
101. Siegler, H. F., Shingleton, W. W., Metzgar, R. S. 1972. Non-specific and specific immunotherapy in patients with melanoma. *Surgery* 72:162–74
102. Milton, G. W., McCarthy, W. H. 1975. Immunotherapy for malignant melanoma. See Ref. 100, p. 251
103. Pinsky, C. M., Hirshout, Y., Wanebo, H. 1976. Randomized trial of Bacillus Calmette Guerin (percutaneous administration) as surgical adjuvant. Immunotherapy for patients with stage II melanoma. *Ann. NY Acad. Sci.* 277:187
104. Laucius, J. F., Bodurtha, A. J., Berkelhamer, J., Prehn, L. M., Prehn, R. T., Mastrangelo, M. J. 1974. Intradermal injection of autochthonous irradiated tumor cells plus BCG in the treatment of metastatic malignant melanoma. *Proc. Am. Assoc. Cancer Res.* 15:171
105. Currie, G. A., McElwain, T. J. 1975. Active immunotherapy as an adjunct to chemotherapy in the treatment of disseminated malignant melanoma: A pilot study. *Br. J. Cancer* 31:143–56
106. Minton, J. P. 1973. Mumps virus and BCG vaccine in metastatic melanoma. *Arch. Surg.* 106:503–6
107. Pinsky, C., Hirshaut, Y., Oettgen, H. 1973. Treatment of malignant melanoma by intratumoral injection of BCG. 39:225–28
108. Nathanson, L. 1972. Regression of intradermal malignant melanoma after

intralesional injection of *Mycobacterium bovis* strain BCG. *Cancer Chemother. Rep.* 56:659–65

109. Mastrangelo, M. J., Kimm, Y. H., Bornstein, R. S. 1974. Clinical and histologic correlation of melanoma regression after intralesional BCG therapy: A case report. *J. Natl. Cancer Inst.* 52:19–24
110. Yamamura, Y., Yoshizaki, K., Azuma, I., Yagura, T., Watanabe, T. 1975. Immunotherapy of human malignant melanoma with oil-attached BCG cell-wall skeleton. *Gann* 66:355–63
111. Richman, S. P., Gutterman, J. U., Hersh, E. M., in preparation
112. Everall, J. D., Wand, J., O'Doherty, C. J., Dowd, P. M. 1975. Treatment of primary melanoma by intralesional vaccinia before excision. *Lancet* 2:583–86
113. Mavligit, G. M., Gutterman, J. U., Burgess, M. A., Khankhanian, N., Seibert, B., Speer, J. F., Jubert, A. V., Martin, R. C., McBride, C. M., Copeland, E. M., Gehan, E. A., Hersh, E. M. 1976. Prolongation of post operative disease-free interval and surgery in human colorectal cancer by Bacillus Calmette-Guerin (BCG) or BCG plus 5-fluorouracil. *Lancet* 1:871–75
114. Moertel, C. G., Ritts, R. E., Schutt, A. J., Hahn, R. G. 1975. A phase I study of MER (methanol extraction residue) fraction of BCG as an immunostimulant in patients with advanced cancer. *Cancer Res.* 35:3057
115. Gottlieb, J. A., Blumenschein, G. R., Gutterman, J. U. 1975. Adriamycin in the treatment of breast cancer. *Adriamycin Rev.* 4:249
116. Gutterman, J. U., Blumenschein, G. R., Hortobagyi, G., Mavligit, G. M., Hersh, E. M. 1976. Immunotherapy for breast cancer. *Breast—Dis Breast* 2:29
117. Donaldson, R. C. 1972. Methotrexate plus Bacillus Calmette-Guerin (BCG) and Isoniazid in the treatment of cancer of the head and neck. *Am. J. Surg.* 124:527
118. Richman, S. P., Livingston, R. B., Gutterman, J. U., Suen, J. Y., Hersh, E. M. 1976. Chemotherapy versus chemoimmunotherapy of head and neck cancer: Report of a randomized study. *Cancer Treat. Rep.* 60:535–39
119. Sokal, J. E., Aungst, W., Snyderman, M. 1974. Delay in progression of malignant lymphoma after BCG vaccination. *N. Engl. J. Med.* 291:1226
120. Mathé, G., Belpomme, D., Pouillart, P., Schwarzenberg, L., Misset, J. L., Jasmin, M. C., Musset, M., Cattan, A., Amiel, J. L., Schneider, M. 1975. Preliminary results of an immunotherapy trial on terminal leukaemic lymphosarcoma. *Lett. Biomed.* 23:465–67
121. Townsend, C. M., Eilber, F. R., Morton, D. L. 1976. Skeletal and soft tissue sarcomas: Results of treatment with adjuvant immunotherapy. *J. Am. Med. Assoc.* 236:2187–89
122. Eidinger, D. and Morales, A. 1976. Treatment of superficial bladder cancer in man. In press
123. Weiss, D. W. 1976. MER and other Mycobacterial fractions in the immunotherapy of cancer. *Med. Clin. North Am.* 60:473–97
124. Hortobagyi, G., in preparation
125. Ribi, E., Granger, D. L., Milner, K. C., Strain, S. M. 1975. Brief communication: tumor regression caused by endotoxins and mycobacterial reactions. *J. Natl. Cancer Inst.* 55:1253–57

Ann. Rev. Med. 1977. 28:517–28
Copyright © 1977 by Annual Reviews Inc. All rights reserved

APPLICATIONS AND HAZARDS OF INTRAVENOUS HYPERALIMENTATION

❖7258

Stanley J. Dudrick, M.D. and James M. Long, III, M.D.
Department of Surgery, The University of Texas Medical School at Houston,
and Hermann Hospital, Houston, Texas 77030

INTRODUCTION

Among the primary functions of the gastrointestinal tract are ingestion, digestion, and absorption of food and assimilation of resulting substrates into the body cell mass. Any one or more of a variety of pathologic processes may significantly compromise any or all of these vital functions, and homeostasis and nutritional integrity might well become dependent upon parenteral administration of essential biochemical substrates for extended periods of time. The successful clinical application of intravenous hyperalimentation has allowed total parenteral nutrition to be a reality for the first time (1). Using this technique, it is now possible to provide the quantity and quality of essential nutrients necessary to maintain positive nitrogen balance and to achieve weight stability or gain in adults and normal growth and development in infants under a wide variety of conditions usually associated with a catabolic response. Having the capacity to feed patients adequately by vein has given the physician an option to bypass a malfunctioning alimentary tract for as long as necessary to allow return of normal or maximal function, while simultaneously maintaining or improving the nutritional status of the patient. Moreover, a period of relative gastrointestinal-tract rest, which total parental nutrition makes possible, allows time for the diseased or disordered segment of the system to be spared from secretory and mechanical stimulation and its usual functions while it is being repaired or restored. Furthermore, the feasibility of administering all required nutrient substrates by parenteral means has provided basic and clinical investigators with a valuable and novel model for studying the physiologic and pathophysiologic processes and reactions of the gastrointestinal tract and other body systems and for achieving absolute control of substrate input for metabolic studies.

Restoration and maintenance of nutritional and metabolic homeostasis must be concomitant goals throughout all phases of diagnosis, therapy, recovery, convalescence, and rehabilitation of patients (2). This is best accomplished by conscientious attention to dietary regimens, tailored to the specific requirements and conditions in individual patients, and delivered to the gastrointestinal tract by mouth, by a variety of enteral feeding tubes, or by surgical ostomies whenever possible and practical. However, when use of the alimentary tract for delivering nutrients is inadequate, ill-advised, or impossible for extended periods of time, nourishment must be supplemented or accomplished entirely by parenteral routes. The basic indications for application of total parenteral nutrition clinically are for support and management of patients who cannot eat, will not eat, should not eat, or cannot eat enough.

Applications when Adequate Enteral Intake is Impossible

The many specific conditions in which use of the alimentary tract for optimally nourishing a patient may be impossible can be classified into the following categories: (*a*) mechanical obstruction; (*b*) reflex or postoperative ileus; (*c*) acute inflammatory, catabolic, or malabsorptive processes; and (*d*) short-gut syndrome.

Enteral nutrition is impaired by obstruction of the alimentary tract by any lesion or disorder that either totally or partially impedes the natural flow of nutrients through the lumen. Conditions that gradually produce obstruction of the alimentary tract often cause a progressive and insidious form of malnutrition that may be easily corrected by total parenteral nutrition. However, sudden onset of acute obstruction of the alimentary tract usually requires prompt operative intervention before a meaningful program of nutritional rehabilitation can be undertaken. The most frequent and significant complications following definitive operations for lesions that have gradually obstructed the upper gastrointestinal tract are related directly to the nutritional deficits incurred before operation. Every patient who is to undergo a surgical procedure for a nonacute obstructing lesion of the alimentary tract should be given serious consideration as a candidate for preoperative parenteral feeding. Thus, by using parenteral nutrition, not only can the incidence of postoperative anastomotic failure be reduced, but the patient can better tolerate the effects of the catabolic response to the operation and avoid many of the other complications that often prolong morbidity or increase mortality. If an anastomotic leak occurs, it can frequently be managed satisfactorily by a conservative regimen of simple drainage, bowel rest, and total parenteral nutrition, thus avoiding another major operation in most cases.

Complete obstruction of the small or large bowel is usually treated best by immediate operation. Occasionally, when the question of complete versus partial obstruction is unanswered, especially when the primary etiology is regional enteritis, granulomatous colitis, ulcerative colitis, or diverticulitis, patients can be partially replenished with vital biochemical substrates until the diagnosis is established and definitive therapy is undertaken. Operation should not be delayed when the clinical condition of the patient suggests imminent danger. On the other hand, every opportunity should be taken to supply the depleted cellular milieu with essential nutrients from which life-sustaining proteins can be synthesized.

Total parenteral nutrition in children has been particularly applicable to situations in which extensive operative resection or reconstruction of gastrointestinal continuity has been necessary during the neonatal period, as well as to conditions that render the alimentary tract temporarily ineffective as a means of sustaining normal nutrition. Despite the gravity of the insult imposed by lesions of the alimentary tract, infants have been successfully supported during preoperative, operative, and postoperative management of complex congenital anomalies. Moreover, intravenous hyperalimentation allows growth and development of the infant, permitting specific organs to mature, heal, or recover sufficiently to achieve normal function or to adapt to resected or altered components of the alimentary tract. The problem of managing severely premature or immature neonates is just as great as managing an infant with the short-gut syndrome. In both instances, the infant's inability to maintain normal enteral nutrition is aggravated by limited organ function. Total parenteral nutrition can support normal growth and development of the infant until the alimentary tract is functional and self-sufficient.

Prolonged adynamic ileus may result from myriad metabolic, traumatic, reflex, or inflammatory disorders. In all of these conditions, the association of starvation with the catabolic process, determined in large measure by the etiology of the ileus, establishes a sequence of events that can lead to steady deterioration and death if the process is not interrupted. If a patient is unable to resume oral intake within seven days after an operation, parenteral nutrition should be instituted. However, total parenteral nutrition is never a substitute for adequate surgical or medical therapy, and abscesses must be drained, obstructions relieved, foreign bodies removed, infections controlled, and fluid and electrolyte balance achieved for optimal patient management to be realized.

Enteral feeding is not only frequently impossible, but may be contraindicated in a patient who has acute pancreatitis. Bowel rest appears to be one of the most important therapeutic measures in managing this disorder because of the reduced stimulation of digestive hormones and enzymes. However, total bowel rest imposes de facto a state of starvation upon a patient who already has increased demands for nutrients. The application of total parenteral nutrition has obviated this therapeutic dilemma. Moreover, pancreatic fistulas have closed spontaneously while the patient was maintained on this regimen (3).

In patients who have sustained major full-thickness burns or other complicated traumas, and in whom nutrient requirements cannot be met by enteral feeding, as many as 5000 to 7000 cal/day have been provided totally by vein or by using a combination of oral, tubular and intravenous feedings. Using combination techniques, over 200 g of protein equivalent have been given to selected patients with multiple injuries.

Massive loss of small bowel, for whatever reason, leaves the patient in a situation in which normal oral nutritional sustenance is impossible. Following massive resections, the residual bowel is incapable of immediately performing the secretory and absorptive functions formerly accomplished by the entire intestine. Applying the technique of total parenteral nutrition for several months in such patients has been lifesaving (4).

Applications when Adequate Enteral Intake is Inadvisable

During the early phases of granulomatous enterocolitis, mucosal involvement passes through a phase that may be compared to a first-degree burn, that is, minimal inflammation without significant loss of tissue and with the possibility of complete recovery. Another group consists of patients with granulomatous enterocolitis, who have bowel injuries roughly comparable to second-degree burns, that is, serious epithelial damage from which recovery is possible if the residual viable basilar cells are adequately protected from further injury. The institution of a regimen of bowel rest and total parenteral nutrition has significant application in both of these circumstances. In some patients, however, the inflammatory response is so severe that spontaneous complete recovery of the bowel is unlikely or impossible, a condition analogous to a third-degree burn. Under these circumstances, operation is usually required.

In a series of 52 consecutive patients with inflammatory bowel disease treated with intravenous hyperalimentation, patients with granulomatous enterocolitis responded more favorably than those with chronic or acute ulcerative colitis (4–6). The responses to bowel rest and total parenteral nutrition delineated three categories to which the burn analogy could again be applied: (*a*) those patients in whom the disease became quiescent in a few weeks and normal oral intake was reinstituted, a situation comparable to first-degree burns; (*b*) those in whom the disease became inactive after several weeks of intravenous hyperalimentation, a group analogous to second-degree burns; and (*c*) those in whom the disease process did not respond and in whom some form of operative procedure was required, similar to third-degree burns. The second group could actually be subdivided into two categories analogous to superficial areas that usually heal completely, and more deeply diseased areas that do not heal and often become the foci of abscesses, fistulas, or obstructions, as a result of superimposed infection and mechanical or chemical trauma.

In our clinical study, 53.8% of the patients responded to bowel rest and total parenteral nutrition by sufficient healing of the involved tissues to allow resumption of normal oral feeding. During the period of intravenous hyperalimentation, which averaged 36.3 days for each patient in this group, the epithelium apparently healed in a manner similar to that of a first-degree burn. The second group consisted of 13 patients (25%) in whom remission of the disease was achieved, but operation was required because of a complication associated with some other manifestation of the disease process; they were analogous to second-degree burn patients. Eleven of the patients (21.2%) had a favorable response to the regimen of bowel rest and total parenteral nutrition, but did not show any objective evidence of improvement of the underlying disease process. Although all of these patients eventually required operation, many of them did not have evidence of cicatrix formation and obstruction. However, the damage to the bowel was sufficiently advanced that even with the prolonged elimination of intraluminal food and reduction of digestive juices from the diseased bowel segments, appreciable mucosal regeneration had not occurred. These patients were analogous to the third-degree burn patients.

In all categories of inflammatory bowel disease treated by a program of total parenteral nutrition, significant improvement of the overall nutritional status of the patient was observed, regardless of the response of the disease. Even when operation was ultimately required, the patient was much better able to tolerate the surgical trauma and to achieve prompt and complete recovery and rehabilitation. We conclude, therefore, that total parenteral nutrition is indicated in patients who have granulomatous enterocolitis in each of the following conditions: (*a*) prolonged periods of active disease associated with significant catabolism; (*b*) preoperative and postoperative periods of patients requiring operation; (*c*) intestinal obstruction; (*d*) short-bowel syndrome resulting from massive intestinal resection; (*e*) indolent wounds and fistulas; and (*f*) periods of bowel rest during acute exacerbation of the disease.

Prior to the clinical application of intravenous hyperalimentation, the management of internal and external intestinal fistulas was accompanied by significant morbidity and very high mortality rates. Virtually all patients with such complications suffered some degree of nutritional depletion, which was further aggravated by presence of a fistula. Although the nutritional deficits were obvious, a uniformly successful technique for the repair of such deficits was not available. The level of the fistula within the alimentary tract usually determined whether or not some form of enteral alimentation could be accomplished. A fistula from the upper intestinal tract, such as the duodenum or jejunum, could not be managed successfully in most instances with enteral feedings. Fistulas from the distal portions of the alimentary tract, particularly the colon, were more likely to close during oral feeding. For these reasons, prompt operation was the classical treatment for all upper gastrointestinal-tract fistulas; the operative risk was high, and failure was common. Application of the technique of bowel rest and total parenteral nutrition has eliminated the urgency for early high-risk operations and has yielded much more satisfactory results (7).

Our series now includes more than 100 consecutive patients with enterocutaneous or enteroenteral fistulas, and we have achieved a spontaneous closure rate of 70.5% of those fistulas. The overall mortality rate in this group has been 6.45%, which is in contrast to previously reported mortality rates of 40–65% in similar series. Of the 21.8% of our patients who required operation, the mortality rate was only 6.67%. The 7.7% of fistulas that never did close were all in patients who died.

Several specific conditions can inhibit spontaneous closure of internal and external fistulas despite an adequate program of bowel rest and total parenteral nutrition. Distal bowel obstruction, the presence of a foreign body, epithelialization of the fistula tract, the presence of an adjacent intraabdominal abscess, and sometimes the effects of irradiation or of contiguous malignant disease will render nonoperative therapy inadequate. Recently, however, spontaneous closure has been achieved in some patients who had fistulas associated with irradiation damage or neoplasm. Total parenteral nutrition serves as an excellent means for preparing such patients for operation, but spontaneous fistula closure should not be expected or sought when such major obstacles are present.

Patients who have acute or chronic renal failure and in whom feeding via the alimentary tract is either impossible or inadequate have also been supported satisfactorily by vein with specially formulated solutions of L-amino acids, hypertonic dextrose, vitamins, and minerals (8, 9). In addition to the usual catabolic state in such patients, a self-imposed state of semi-starvation often results from the associated anorexia. In other patients who have some degree of hepatic dysfunction, total parenteral feeding has been successful by using diets tailored to specific amino acid requirements that appear to exist under these conditions (10). Finally, patients with certain cardiac disorders have shown significant improvement both in myocardial function and nutritional status following relatively short periods of parenteral nutritional support (11).

Applications for which Adequate Enteral Intake is Improbable

Table 1 lists a diverse group of disease states in which the maintenance of a balanced enteral dietary regimen is very unlikely, even though the alimentary tract may be potentially or partially functional. Some of these conditions appear more than once in the table because the disease process may have very different effects on enteral nutrition in particular patients. The interrelationships of these conditions to each other and to malnutrition in general need no further comment.

The nutritional status of patients receiving cancer chemotherapy warrants brief discussion. Many of these patients are already in a state of significant malnutrition secondary to a combination of the parasitic nature of the malignant process itself, the catabolism of previous surgical procedures or postoperative complications, and the anorexia associated with both. Chemotherapeutic agents themselves cause additional insult to the nutritional status of the patient by aggravating catabolic processes and often rendering the patient anorectic. In cancer patients receiving chemotherapy, radiation therapy, immunotherapy, or a combination of modalities, application of bowel rest and total parenteral nutrition has reduced the manifestations of alimentary-tract toxicity and has allowed administration of as much as two to three times the usual chemotherapeutic dose without serious deleterious effects and with enhanced salutary results in many instances (12).

In many patients who have malignant diseases or other debilitating processes, immunoincompetence, indicated by anergic responses to a battery of skin-test antigens, has been reversed and normal immunologic mechanisms have been restored when sufficient nutritional support was provided via intravenous hyperalimentation (13, 14). Early results indicate that this effect improves the ability of the patient to resist infection and bolsters natural immunologic defense mechanisms against the neoplastic process. The potential of this technique as an adjunct to the various forms of cancer therapy has not yet been fully explored or exploited.

In patients with such disorders as anorexia nervosa, hyperemesis gravidarum, and senile dementia, total parenteral nutrition can be a valuable technique for restoring nutritional status when attempts to use the gastrointestinal tract are ineffective. Finally, maintenance of the anabolic state in patients who have major debilitating disorders or who require major operative procedures may decrease the incidence of complications, reduce the duration and cost of hospitalization, and achieve more rapid and complete patient recovery and rehabilitation.

Applications when Adequate Enteral Intake is Hazardous

Each of the conditions listed in Table 1 has an apparent association with malnutrition, but to use standard diets or even specialized enteral feeding techniques can be hazardous and potentially lethal. In almost all of these situations, oral or tubular feedings are associated with an unacceptably high incidence of aspiration pneumonia. When properly carried out, total parenteral nutrition affords the safest and most judicious means of achieving satisfactory metabolic rehabilitation of such patients.

Table 1 Applications of total parenteral nutrition

To avoid or correct malnutrition when adequate enteral intake is:			
Impossible	Improbable	Inadvisable	Hazardous
1. Alimentary-tract obstruction or ileus a. Oropharynx—neoplasm b. Esophagus—achalasia, neoplasm, stricture c. Stomach—benign or malignant ulcer or neoplasm, postoperative atony, gastroschisis d. Duodenum—ulcer, neoplasm, veil, atresia, duplication, choledochal cyst, annular pancreas, carcinoma of the head of the pancreas, superior mesenteric artery syndrome e. Small intestine—atresia, intussusception, neoplasm, polyp, regional enteritis, adhesions, bands, veils, gallstone ileus, foreign body, malrotation f. Colon—diverticulitis, volvulus, neoplasm, ulcerative colitis, granulomatous colitis	1. Malnourished geriatric patients	1. Enterocutaneous and enteroenteral fistulas	1. Cerebrovascular accident
2. Acute hemorrhagic pancreatitis	2. Chronic starvation—marasmus, kwashiorkor	2. Renal failure—acute or chronic	2. Mental obtundity
3. Short-gut syndrome—2 months postoperatively	3. Anorexia nervosa	3. Acute hepatitis	3. Reversible coma
4. Radiation effects—entire GI tract	4. Hyperemesis gravidarum	4. Reversible liver failure	4. Tetanus
5. Chemotherapy effects—entire GI tract	5. Severe mental depression	5. Partial intestinal obstruction	5. Laryngeal incompetence
6. Infectious gastroenteritis—cholera, salmonellosis, shigellosis	6. Chronic alcoholism	6. Acute pancreatitis	6. Tracheoesophageal fistula
7. Intraabdominal abscess and/or peritonitis	7. Chronic relapsing pancreatitis	7. Acute regional enteritis	7. Toxic megacolon
8. Wound dehiscence	8. Recurrent regional enteritis	8. Acute granulomatous or ulcerative colitis	8. Meteorism
9. Protein-losing gastroenteropathy—sprue, Menétrièr's disease	9. Chronic granulomatous or ulcerative colitis	9. Chronic respirator dependence	9. Occult gastrointestinal bleeding
10. Massive catabolism—major burns, multiple trauma	10. Radiation effects	10. Severe alkaline reflux gastritis	
11. Other congenital anomalies—omphalocele, diaphragmatic hernia	11. Chemotherapy effects	11. Severe cachexia	
	12. Short-gut syndrome—after maximal adaptation	12. Malabsorption syndromes	
	13. Myocardial insufficiency	13. Extreme prematurity	
	14. Chronic respiratory-tract disease	14. Failure to thrive	
	15. Bronchopleural fistula	15. Intractable diarrhea	
	16. Moderate catabolism—burns, trauma, multiple operations		

Hazards of Total Parenteral Nutrition

The infectious and technical problems related to long-term central venous catheters are listed in Table 2. Of course, prevention of infection and sepsis is of utmost importance to the safety and efficacy of long-term total parenteral nutrition. If aseptic and antiseptic principles are conscientiously practiced during the insertion and subsequent maintenance of the indwelling central venous feeding catheter, the incidence of catheter-related sepsis is indeed very low (15, 16). Because hypertonic nutrient solutions are excellent culture media for many species of bacteria and fungi, meticulous asepsis must also be maintained during solution preparation, additive injection, and long-term infusion. If fever should occur in a patient receiving total parenteral nutrition, the physician must immediately attempt to define its source in the ears, nose or throat, chest, urinary tract, gastrointestinal tract, or wound. If the etiology of the fever cannot be determined readily, the nutrient solution and tubing must be empirically and promptly replaced, and specimens of the blood and the solution must be cultured for aerobic and anaerobic bacteria and for fungi. Should the fever persist after replacement of the solution and administration apparatus, the infusion must be terminated immediately. The indwelling subclavian vein catheter

Table 2 Hazards of total parenteral nutrition related to central venous catheterization

Infectious	Technical
1. Insertion-site contamination	Pneumothorax
a. Contamination during insertion	Tension pneumothorax
b. Contamination during routine care	Hemothorax
2. Catheter contamination	Hydrothorax
a. Improper technique inserting catheter	Hydromediastinum Cardiac tamponade
b. Administration of blood via feeding catheter	Brachial plexus injury Horner's syndrome
c. Use of catheter to measure central venous pressure	Phrenic nerve paralysis Carotid artery injury
d. Use of catheter to obtain blood samples	Subclavian artery injury Subclavian hematoma
e. Use of catheter to administer medications	Thrombosis, subclavian vein or superior vena cava
f. Contaminated solution during preparation or additives	Arteriovenous fistula Venobronchial fistula
g. Contaminated tubing via connections	Air embolism Catheter embolism
h. Three-way stopcocks in system	Thromboembolism
3. Secondary contamination	Catheter misplacement
a. Septicemia, bacterial or fungal	Cardiac perforation
b. Septic emboli	Endocarditis
c. Osteomyelitis of clavicle	Thoracic duct laceration
d. Septic arthritis	Innominate or subclavian vein laceration
e. Endocarditis	

must be removed, and the distal tip of the catheter must be cultured for aerobic and anaerobic bacteria and for fungi. Depending upon the clinical situation, another feeding catheter may be inserted into the opposite subclavian vein, or administration of isotonic dextrose solution may be started via a peripheral vein to insure against postinfusion hypoglycemia. Systemic broad-spectrum antibiotic therapy or antifungal therapy is rarely required. However, if desired, or if the patient shows persistent signs of sepsis, an antimicrobial regimen may be instituted at this time and modified when specific sensitivity testing has been completed.

The existence of fever or evidence of sepsis prior to the institution of total parenteral nutrition does not necessarily contraindicate the use of the technique (17). In a traumatized, debilitated, or critically ill patient, sepsis often accentuates the nutritional requirements. Specific antimicrobial therapy has already been instituted in most of such patients, and although seeding of the indwelling catheter by circulating microorganisms is a definite possibility, it has not proved to be an overwhelming clinical problem. In fact, resolution of systemic infections has occurred regularly during the periods of total parenteral nutrition. However, if the clinician suspects that the infectious course of the patient might be caused or aggravated by the superior vena caval catheter, it should be removed and cultured promptly.

Although development of thrombophlebitis is always possible with the use of long-term indwelling catheters and hypertonic solutions, superior vena caval thrombosis has not been observed clinically in more than 4000 patients. Apparently, the high rate of blood flow in this large-diameter vessel assures prompt dilution of the hypertonic fluid and prevents chemical phlebitis. Furthermore, attention to sterility has practically eliminated infectious thrombophlebitis. However, rare incidences of thrombosis have occurred in patients when the catheter tips were misdirected into or malpositioned in an external jugular, internal jugular, internal mammary, or axillary vein. Other central venous catheter–related hazards, such as an inadvertent air embolism and catheter embolism, can be avoided easily by adherence to principles and techniques of safe central venous catheterization reported in previous publications (1, 18, 19). A thorough knowledge of anatomy, combined with common sense and strict adherence to the principles and techniques of percutaneous subclavian vein catheterization, should minimize the potential hazards of accidental pneumothorax, hydrothorax, hemothorax, hydromediastinum, subclavian artery puncture, bleeding, arteriovenous fistula, endocarditis, injury to the thoracic duct, venobronchial fistula, injury to the brachial plexus or sympathetic chain, osteomyelitis of the clavicle, or even death.

Nonketotic hyperosmolar hyperglycemia can be precipitated acutely by infusion of the hypertonic nutrient solution too rapidly with resultant marked osmotic diuresis, serum and urine electrolyte aberrations, dehydration, and central nervous system derangements. This syndrome can also occur insidiously when chronically impaired glucose utilization is not recognized, particularly in the patient with diabetes mellitus, sepsis, extensive burns, major trauma, or intracranial operations. If blood and urine sugar concentrations are not conscientiously monitored in such patients, significant hyperglycemia and dehydration can ensue, with accompanying weakness, listlessness, and eventually coma. Basic treatment of nonketotic hyperos-

Table 3 Potential metabolic hazards of total parenteral nutrition

Hazards	Possible explanations
1. Glucose metabolism:	
a. Hyperglycemia, glycosuria, osmotic diuresis, nonketotic hyperosmolar dehydration and coma	Excessive total dose or rate of infusion of dextrose; inadequate endogenous insulin; glucocorticoids; sepsis
b. Ketoacidosis of diabetes mellitus	Inadequate endogenous insulin response; inadequate exogenous insulin therapy
c. Postinfusion (rebound) hypoglycemia	Persistence of endogenous insulin production secondary to prolonged stimulation of islet cells by high-carbohydrate infusion
2. Amino acid metabolism:	
a. Hyperchloremic metabolic acidosis	Excessive chloride and monohydrochloride content of crystalline amino acid solutions
b. Serum amino acid imbalance	Unphysiologic amino acid profile of the nutrient solution; different amino acid utilization with various disorders
c. Hyperammonemia	Excessive ammonia in protein hydrolysate solutions; deficiencies of arginine, ornithine, aspartic acid, and/or glutamic acid in crystalline amino acid solutions; primary hepatic disorder
d. Prerenal azotemia	Excessive total dose or rate of infusion of protein hydrolysate or amino acid solution
3. Calcium and phosphorus metabolism:	
a. Hypophosphatemia 1. decreased erythrocyte 2,3-diphosphoglycerate 2. increased affinity of hemoglobin for oxygen 3. aberrations of erythrocyte intermediary metabolites	Inadequate phosphorus administration, redistribution of serum phosphorus into cells and/or bone
b. Hypocalcemia	Inadequate calcium administration; reciprocal response to phosphorus repletion without simultaneous calcium infusion; hypoalbuminemia
c. Hypercalcemia	Excessive calcium administration with or without high doses of albumin; excessive vitamin D administration
d. Vitamin D deficiency; hypervitaminosis D	Inadequate or excessive vitamin D administration
4. Essential fatty acid metabolism: Serum deficiencies of phospholipid linoleic and/or arachidonic acids, serum elevations of Δ-5,8,11-eicosatrienoic acid	Inadequate essential fatty acid administration; inadequate vitamin E administration

Table 3 *(Continued)*

Hazards	Possible explanations
5. Miscellaneous:	
a. Hypokalemia	Inadequate potassium intake relative to increased requirements for protein anabolism; diuresis
b. Hyperkalemia	Excessive potassium administration, especially in metabolic acidosis; renal decompensation
c. Hypomagnesemia	Inadequate magnesium administration relative to increased requirements for protein anabolism and glucose metabolism
d. Hypermagnesemia	Excessive magnesium administration; renal decompensation
e. Anemia	Iron deficiency; folic acid deficiency; vitamin B-12 deficiency; copper deficiency; other deficiencies
f. Bleeding	Vitamin K deficiency
g. Hypervitaminosis A	Excessive vitamin A administration
h. Elevations in SGOT, SGPT, and serum alkaline phosphatase	Enzyme induction secondary to amino acid imbalance; excessive glycogen and/or fat deposition in the liver
i. Cholestatic hepatitis	Decreased water content of bile

molar hyperglycemia consists of judicious infusion of isotonic or half-strength solutions of saline or dextrose, together with insulin, while simultaneously obtaining frequent measurements of fluid loss, central venous pressure, serum and urine electrolytes, and blood glucose. Thorough assessment and understanding of the patient's underlying disease processes and metabolic status, and proper use of established principles of the technique of intravenous hyperalimentation should obviate the vast majority of the potential metabolic hazards of total parenteral nutrition listed in Table 3 (20).

Other rare metabolic hazards that can arise during total parenteral nutrition are usually iatrogenic, and include fluid overload, hypernatremia, hyponatremia, hypermagnesemia, trace-element deficiencies, acne, and acalculous cholecystitis. Although these infrequent hazards can usually be avoided by adherence to the established principles and techniques of total parenteral nutrition, the clinician must be ever alert to the possibility of their occurrence.

CONCLUSIONS

The current technique of total parenteral nutrition (intravenous hyperalimentation) is relatively safe and efficacious, but will undoubtedly undergo change and modification as improved techniques and materials become available. With judicious application of the technique as a clinical therapeutic and investigative tool, new indications for general and specific total parenteral nutrition will likely expand and extend into every medical discipline. As relatively few of the potential applications for total parenteral nutrition have been thoroughly explored or exploited, the clinical and laboratory research possibilities appear to be virtually limitless. Thus, as increasing knowledge and experience are gained, applications of total parenteral nutrition will inevitably be expanded and will become more specific, and the hazards will no doubt decrease.

Literature Cited

1. Dudrick, S. J., Wilmore, D. W., Vars, H. M., Rhoads, J. E. 1969. Can intravenous feeding as the sole means of nutrition support growth in the child and restore weight loss in an adult? An affirmative answer. *Ann. Surg.* 169:974–84
2. Artz, C. P., Hardy, J. D., eds. 1975. *Management of Surgical Complications.* Philadelphia: Saunders, 834 pp.
3. Dudrick, S. J., Wilmore, D. W., Steiger, E., Mackie, J. A., Fitts, W. T. 1970. Spontaneous closure of traumatic pancreatoduodenal fistulas with total intravenous nutrition. *J. Trauma* 10: 542–53
4. Bockus, H. L., ed. 1975. *Gastroenterology,* Vol. 2. Philadelphia: Saunders. 1171 pp.
5. Anderson, D. L., Boyce, H. W. 1973. Use of parenteral nutrition in treatment of advanced regional enteritis. *Dig. Dis.* 18:633
6. Clearfield, H. R., Dinoso, V. P., eds. 1976. *Gastrointestinal Emergencies.* New York: Grune & Stratton. 305 pp.
7. MacFadyen, B. V. Jr., Dudrick, S. J., Ruberg, R. L. 1973. The management of gastrointestinal fistulas with parenteral hyperalimentation. *Surgery* 74: 100–5
8. Dudrick, S. J., Steiger, E., Long, J. M. 1970. Renal failure in surgical patients. Treatment with intravenous essential amino acids and hypertonic glucose. *Surgery* 68:180–86
9. Abel, R. M., Beck, C. H. Jr., Abbott, W. M., Ryan, J. A. Jr., Barnett, G. O., Fischer, J. E. 1973. Improved survival from acute renal failure after treatment with intravenous essential L-amino acids and glucose. *N. Engl. J. Med.* 288:695–99
10. Fischer, J. E. 1976. *Total Parenteral Nutrition.* Boston: Little, Brown. 454 pp.
11. Abel, R. M., Fischer, J. E., Buckley, M. J., Austen, W. G. 1974. Hyperalimentation in cardiac surgery. *J. Thorac. Cardiovasc. Surg.* 67:294–300
12. Copeland, E. M., MacFadyen, B. V. Jr., Lanzotti, V. J., Dudrick, S. J. 1975. Intravenous hyperalimentation as an adjunct to cancer chemotherapy. *Am. J. Surg.* 129:167–73
13. Law, D. K., Dudrick, S. J., Abdou, N. I. 1973. Immunocompetence of patients with protein-calorie malnutrition. *Ann. Intern. Med.* 79:545–50
14. Copeland, E. M., MacFadyen, B. V. Jr., Dudrick, S. J. 1976. Effect of intravenous hyperalimentation on established delayed hypersensitivity in the cancer patient. *Ann. Surg.* 184:60–64
15. Copeland, E. M., MacFadyen, B. V. Jr., Dudrick, S. J. 1974. Prevention of microbial catheter contamination in patients receiving parenteral hyperalimentation. *South. Med. J.* 67:303–6
16. Ryan, J. A., Abel, R. M., Abbott, W. M., Hopkins, C. C., Chesney, T. M., Colley, R., Phillips, K., Fischer, J. E. 1974. Catheter complications in total parenteral nutrition. *N. Engl. J. Med.* 290:757–61
17. Copeland, E. M., MacFadyen, B. V. Jr., McGown, C., Dudrick, S. J. 1974. The use of hyperalimentation in patients with potential sepsis. *Surg. Gynecol. Obstet.* 138:377–80
18. Ballinger, W. F., Collins, J. A., Drucker, W. R., Dudrick, S. J., Zeppa, R., eds. 1975. *Manual of Surgical Nutrition.* Philadelphia: Saunders. 527 pp.
19. Nyhus, L. M., ed. 1973. *Surgery Annual.* New York: Appleton-Century-Crofts. 418 pp.
20. Dudrick, S. J., MacFadyen, B. V. Jr., Van Buren, C. T., Ruberg, R. L., Maynard, A. T. 1972. Parenteral hyperalimentation: metabolic problems and solutions. *Ann. Surg.* 176:259–64

CUMULATIVE INDEXES

CONTRIBUTING AUTHORS VOLUMES 24-28

A

Abelmann, W. H., 24:145-52
Aberdeen, E., 26:451-64
Ahmed, S. S., 28:393-409
Ahn, Y. S., 28:299-309
Alexander, P., 27:207-24
Anthony, B. F., 28:355-69
Arnold, A. B., 28:187-94
Arnone, A., 25:123-30
Atkinson, A. J., 25:99-113
Atkinson, J. P., 28:291-98

B

Bagshaw, M. A., 26:567-88
Baldwin, D. S., 27:49-55
Baldwin, R. W., 27:151-63
Barchas, J. D., 26:499-520
Barker, W. H. Jr., 25:75-81
Barlow, G. H., 24:335-44
Barnes, A. C., 27:1-7
Becker, M. A., 25:15-28
Begemann, H., 28:131-41
Beisel, W. R., 26:9-20
Berman, M. L., 26:391-401
Bernard, J., 25:39-51
Betts, R. F., 28:103-10
Biggar, W. D., 24:135-44
Blum, C. B., 26:365-71
Bone, H. G. III, 28:111-17
Boshes, B., 26:465-70
Bouhuys, A., 24:111-34
Braverman, L. E., 26:443-49
Brenner, B. M., 28:91-102
Breslow, J. L., 24:315-24
Brewer, G. J., 25:29-38
Brummer, D. L., 25:115-22
Brunner, E. A., 26:391-401
Buckner, C. D., 25:179-98
Burke, J. F., 24:289-94

C

Cadoret, R. J., 26:21-25
Cannon, P. J., 27:485-521
Canterbury, J. M., 24:217-32
Caro, W. A., 28:251-69
Cavalieri, R. R., 28:57-65
Chase, T. N., 26:419-26
Cheng, S. C., 26:391-401
Coburn, R. F., 24:241-50
Cohen, S., 26:373-90
Collen, D., 25:447-55
Collinsworth, K. A., 25:143-48
Conklin, R. H., 28:167-77
Craig, W. A., 27:287-300
Creger, W. P., 24:75-82
Cuatrecasas, P., 24:233-40

D

Dale, D. C., 26:277-81
Davison, R., 25:99-113
Dean, R. H., 27:397-405
Deftos, L. J., 25:323-31
DeGraff, A. C. Jr., 24:111-34
DeLuca, H. F., 25:349-67
Dement, W. C., 27:465-84
Desbuquois, B., 24:233-40
DeVries, W. C., 26:203-17
Dixon, F. J., 25:83-98
Doherty, J. E., 26:159-71
Downing, D. T., 26:27-32
Dudrick, S. J., 28:517-28
DuPont, H. L., 24:103-10; 28:167-77
Durie, B. G. M., 26:283-88
Durrer, D., 27:63-68

E

Edgington, T. S., 25:303-14
Ehrenkrantz, N. J., 25:9-14
Eickhoff, T. C., 28:411-23
Elin, R. J., 27:127-41
Elliott, G. R., 26:499-520
Eriksson, M., 24:29-40
Ettinger, P. O., 28:393-409
Evenson, M. A., 28:85-90

F

Fanelli, G. M. Jr., 28:349-54
Fauci, A. S., 26:277-81
Fefer, A., 25:179-98
Feldman, D., 26:83-90
Fernstrom, J. D., 25:1-8
Field, J. B., 24:309-14
Fisher, R. S., 26:373-90
Forget, B. G., 26:345-51
Formal, S. B., 24:103-10
Forstot, S. L., 28:21-35
Frank, M. M., 28:291-98
Fredrickson, D. S., 24:315-24
Freedman, D. X., 24:153-64
Freeman, R. G., 27:77-87
Friedman, R. C., 27:57-62
Friesen, H., 24:251-70
Froelicher, V. F. Jr., 28:1-12
Fudenberg, H. H., 24:175-208

G

Gabbay, K. H., 26:521-36
Gadek, J., 28:291-98
Gangarosa, E. J., 25:75-81
Gault, M. H., 27:117-25
Gavras, H., 27:485-521
Gerich, J. E., 27:379-88
Gibbons, R. J., 26:121-36
Glenner, G. G., 25:131-35
Gluck, F. W., 27:397-405
Glucksberg, H., 25:179-98
Goldberg, M., 26:537-50
Goldreyer, B. N., 26:219-28
Goldsmith, R. S., 27:181-90
Good, R. A., 24:135-44
Gottlieb, L. S., 25:411-29
Grams, R. R., 27:199-206
Green, R., 27:57-62
Greenblatt, D. J., 27:407-20
Grossman, M., 28:43-55
Grumet, F. C., 28:425-52
Guillemin, R., 27:379-88
Guilleminault, C., 27:465-84
Gunnar, R. M., 26:471-83
Gunnells, J. C. Jr., 26:259-75
Gutterman, J. U., 28:489-515

H

Hackett, T. P., 27:371-78
Haider, B., 28:393-409
Hanshaw, J. B., 28:103-10
Harker, L. A., 25:383-400
Harmel, M. H., 26:137-41
Harrington, W. J., 28:299-309
Harrison, D. C., 25:143-48
Hartmann, R. C., 28:187-94
Havel, R. J., 28:195-209
Henderson, E. S., 26:289-306
Henkin, R. E., 26:193-201
Herndon, J. H. Jr., 27:77-87
Hersh, E. M., 28:489-515
Higby, D. J., 26:289-306
Hinman, F., 24:83-88
Hirschhorn, K., 24:67-74
Holick, M. F., 25:349-67
Holland, J. F., 28:475-88
Hollister, L. E., 27:431-46
Holman, R. B., 26:499-520
Holmberg, L., 26:33-44
Hook, E. W., 28:371-91
Hopewell, P. C., 27:343-56

Hornick, R. B., 24:103-10
Hough, P. V. C., 25:431-46
Hurst, M. W., 27:301-12
Hwang, P., 24:251-70

I

Ingbar, S. H., 26:443-49

J

Jacquillat, C., 25:39-51
Janowitz, H. D., 27:269-85
Jenkins, C. D., 27:301-12
Johannes, R. S., 26:589-601
Johnson, J. W. C. Jr., 24:165-74
Johnson, M. T., 25:251-67
Jones, S. R., 25:251-67
Jordan, P. H. Jr., 27:191-97

K

Kane, J. J., 26:159-71
Kaplan, B. S., 27:117-25
Kattus, A. A., 27:69-75
Kaufman, H. E., 28:21-35
Kelly, J. F., 25:53-68
Kennedy, B. J., 26:235-43
Kessler, S., 24:89-102
Klassen, J., 27:117-25
Klein, M. S., 27:89-108
Klein, P. A., 28:311-27
Klerman, G. L., 25:199-217
Koffler, D., 25:149-64
Korenman, S. G., 27:357-70
Kunin, C. M., 27:287-300
Kwaan, H. C., 24:335-44

L

Langer, G. A., 28:13-20
Lanning, C. F., 26:137-41
Lasseter, K. C., 25:69-74
Le Quesne, L. P., 26:63-74
Lerner, K. G., 25:179-98
Levey, G. S., 25:69-74
Levin, A. S., 24:175-208
Levin, N. W., 24:365-84
Levy, R. I., 26:365-71
Lewy, P. R., 24:365-84
Linnemann, C. C. Jr., 28:179-85
Lipinski, J., 26:551-65
Lipscomb, K., 28:271-89
Lofgren, S., 27:407-20
Long, J. M. III, 28:517-28
Lopez, M., 27:453-63
Louria, D. B., 25:219-31
Love, R., 28:131-41
Luby, J. P., 25:251-67
Lyons, M. M., 28:393-409

M

Maddox, D. A., 28:91-102
Marcus, R., 27:357-70
Matthysse, S., 26:551-65
Mavligit, G. M., 28:489-515
McCabe, W. R., 27:335-41
McCarty, D. J., 25:279-88
McConnell, R. B., 25:165-78
McCullough, J. L., 24:345-52
McDevitt, H. O., 28:425-52
McGraw, C. P., 26:321-29
McGuffin, W. L. Jr., 26:259-75
McGuigan, J. E., 25:233-49
McGuire, W. L., 26:353-63
McPhaul, J. J. Jr., 28:37-42
McRee, D. I., 27:109-15
Mello, N. K., 27:321-33
Mendelson, J. H., 27:321-33
Merimee, T. J., 25:137-42
Merin, R. G., 28:75-83
Messner, R. P., 26:181-92
Millman, B. S., 28:223-34
Mitchell, J. H., 25:369-81
Moos, R. H., 24:89-102
Moran, J. F., 26:471-83
Moylan, J. A., 28:85-90
Mullins, C. B., 28:271-89
Murphy, D. L., 24:209-16
Murray, J. F., 27:343-56
Murray, T., 26:537-50

N

Nasenbeny, J., 28:179-85
Nathan, D. G., 26:345-51
Neer, R., 25:323-31
Neill, C. A., 24:61-67
Neiman, P., 25:179-98
Nicoloff, D. M., 26:143-57
Nilsson, I. M., 26:33-44
Norman, P. S., 26:337-44

O

Odell, G. B., 24:165-74
Okada, D. M., 28:355-69
Oken, D. E., 26:307-19
Oldewurtel, H. A., 28:393-409
Oliver, J. A., 27:485-521
Ontjes, D. A., 27:263-68
Oratz, M., 26:91-104
Orci, L., 28:119-30
O'Reilly, R. A., 27:245-61

P

Pak, C. Y. C., 28:111-17
Palmer, R. F., 25:69-74
Parisi, A. F., 28:159-66
Park, B. H., 24:135-44
Patterson, R., 25:53-68
Paulson, G. W., 26:75-81
Perloff, M., 28:475-88
Pierce, A. K., 28:211-21
Pochi, P. E., 26:27-32
Pollack, R. E., 25:431-46
Portnoy, B. L., 28:167-77
Potolsky, A., 24:75-82
Price, M. R., 27:151-63
Prien, E. L. Jr., 26:173-79

Q

Quinn, J. L. III, 26:193-201
Quintanilla, A., 24:365-84

R

Rabinowitz, M., 24:245-62
Rackley, C. E., 26:105-20
Raisfeld, I. H., 24:385-418
Rakowski, H., 27:165-80
Ranganathan, N., 27:165-80
Rapoport, B., 28:57-65
Ray, G. R., 26:567-88
Regan, T. J., 28:393-409
Reiss, E., 24:217-32
Robertson, G. L., 25:315-22
Robertson, J., 28:211-21
Robins, E., 28:67-73
Rose, R. M., 27:301-12
Ross, G. T., 24:295-302
Rosse, W. F., 26:331-36
Rothschild, M. A., 26:91-104
Ruderman, N. B., 26:245-58
Russell, R. O. Jr., 26:105-20
Ryan, W. G., 28:143-52

S

Sachar, D. B., 27:269-85
Sachar, E. J., 27:389-96
Salmon, S. E., 26:283-88
Salvaggio, J., 27:453-63
Sanford, J. P., 25:401-10; 26:485-98; 27:421-29
Sasahara, A. A., 28:159-66
Sasazuki, T., 28:425-52
Schacht, R. G., 27:49-55
Schiff, G. M., 27:447-52
Schildkraut, J. J., 25:333-48
Schreiber, S. S., 26:91-104
Schrier, S. L., 25:269-78
Scott, H. W. Jr., 27:397-405
Seegmiller, J. E., 25:15-28
Senay, E. C., 24:153-64
Sever, J. L., 26:589-601
Shader, R. I., 27:407-20
Shah, P. M., 28:235-50
Sharma, G. V. R. K., 28:159-66
Sharp, G. W. G., 24:19-28
Sherlock, S., 27:143-49
Sherwin, I., 27:37-47
Shoulson, I., 26:419-26

Shull, H. J., 27:397-405
Silen, W., 27:9-22
Silver, M. D., 27:165-80
Sipes, J. N., 28:371-91
Skillman, J. J., 27:9-22
Smith, R. B. III, 26:229-34
Smith, R. T., 28:311-27
Snyder, W. H. III, 28:111-17
Sobel, B. E., 27:89-108
Sokal, J. E., 24:281-88
Sokoloff, L., 24:271-80
Spitler, L. E., 24:175-208
Spitzer, R. L., 27:57-62
Stephen, C. R., 28:153-57
Storb, R., 25:179-98
Strauss, J. S., 26:27-32
Stutzman, L., 24:325-34

T

Terry, W. D., 25:131-35
Theml, H., 28:131-41
Thomas, E. D., 25:179-98
Thompson, R. L., 28:371-91
Tilkian, A., 27:465-84
Tomasi, T. B. Jr., 28:453-65
Toole, J. F., 26:321-29
Tow, D. E., 28:159-66
Trey, C., 25:411-29
Tschopp, T. B., 24:1-18
Twito, D. I., 26:235-43

U

Underwood, L. E., 26:427-41
Unger, R. H., 24:303-8; 28:119-30
Utiger, R. D., 25:289-302

V

Vaitukaitis, J. L., 24:295-302
van Houte, J., 26:131-36
Van Wyk, J. J., 26:427-41
Ventura, A. K., 25:9-14
Vermylen, J., 25:447-55
Verstraete, M., 25:447-55
Voorhees, J. J., 28:467-73

W

Wallace, W. A., 26:1-7
Warren, W. D., 26:229-34
Waterson, A. P., 27:23-35
Wechsler, A. S., 28:341-48
Weil, M., 25:39-51
Weinstein, G. D., 24:345-52
Weitzman, E. D., 27:225-43
Wells, S. A. Jr., 27:263-68
Wender, P. H., 26:45-62
Wessler, S., 27:313-19
Wigle, E. D., 27:165-80
Wilber, J. F., 24:353-64
Wildenthal, K., 25:369-81
Williams, R. C. Jr., 26:181-92
Wilson, C. B., 25:83-98
Wilson, D. R., 28:329-39
Winkelmann, R. K., 28:251-59
Winokur, G., 26:21-25
Wolfe, W. G., 26:203-17
Wolff, S. M., 26:277-81; 27:127-41
Woodrow, J. C., 26:165-78

Y

Yaffe, S. J., 24:29-40
Yen, S. S. C., 26:403-17
Yu, P. N., 26:1-7

Z

Zieve, L., 26:143-57
Zimmerman, T. S., 25:303-14
Zucker, M. B., 24:1-18

CHAPTER TITLES VOLUMES 24-28

ALLERGY AND IMMUNOLOGY		
Immunologic Reconstitution	W. D. Biggar, B. H. Park, R. A. Good	24:135-44
Transfer Factor Therapy in Immune Deficiency States	A. S. Levin, L. E. Spitler, H. H. Fudenberg	24:175-208
Animal Models of the Asthmatic State	R. Patterson, J. F. Kelly	25:53-68
Characterization of Amyloid	G. G. Glenner, W. D. Terry	25:131-35
Immunopathogenesis of Systemic Lupus Erythematosus	D. Koffler	25:149-64
Alterations in T- and B-Cells in Human Disease States	R. C. Williams Jr., R. P. Messner	26:181-92
The Antiglobulin Test in Autoimmune Hemolytic Anemia	W. F. Rosse	26:331-36
Immunotherapy (Densensitization) in Allergic Disease	P. S. Norman	26:337-44
IgA-Associated Glomerulonephritis	J. J. McPhaul Jr.	28:37-42
The Current Status of BCG Immunization Against Tuberculosis	T. C. Eickhoff	28:411-23
The Association of Genes in the Major Histocompatibility Complex and Disease Susceptibility	T. Sasazuki, F. C. Grumet, H. O. McDevitt	28:425-52
Structure and Function of Alpha-Fetoprotein	T. B. Tomasi Jr.	28:453-65
BCG as Adjuvant Immunotherapy for Neoplasia	E. M. Hersh, J. U. Gutterman, G. M. Mavligit	28:489-515
ANESTHESIOLOGY		
Ketamine Anesthesia	C. F. Lanning, M. H. Harmel	26:137-41
Effects of Anesthesia on Intermediary Metabolism	E. A. Brunner, S. C. Cheng, M. L. Berman	26:391-401
Effects of Anesthetic Drugs on Myocardial Performance in Man	R. G. Merin	28:75-83
Malignant Hyperpyrexia	C. R. Stephen	28:153-57
Acupuncture: Context and Critique	B. S. Millman	28:223-34
CARDIOLOGY		
Genetics of Congenital Heart Disease	C. A. Neill	24:61-67
Viral Myocarditis and Its Sequelae	W. H. Abelmann	24:145-52
Primary Hyperlipoproteinemia in Infants	D. S. Fredrickson, J. L. Breslow	24:315-24
Sulfonylureas and the Heart	G. S. Levey, K. C. Lasseter, R. F. Palmer	25:69-74
Diphenylhydantoin as an Antiarrhythmic Drug	A. J. Atkinson, R. Davison	25:99-113
Antiarrhythmic Actions of Lidocaine	D. C. Harrison, K. A. Collinsworth	25:143-48
Static (Isometric) Exercise and the Heart: Physiological and Clinical Considerations	J. H. Mitchell, K. Wildenthal	25:369-81
Sudden Death and the Pre-Hospital Phase of Acute Myocardial Infarction	W. A. Wallace, P. N. Yu	26:1-7
Diagnosis and Prevention of Postoperative Deep-Vein Thrombosis	L. P. Le Quesne	26:63-74
Left Ventricular Function in Acute and Chronic Coronary Artery Disease	C. E. Rackley, R. O. Russell Jr.	26:105-20
Clinical Pharmacology of Digitalis Glycosides	J. E. Doherty, J. J. Kane	26:159-71
Mechanisms of Supraventricular Tachycardias	B. N. Goldreyer	26:219-28
Selective Distal Splenorenal Shunt for Bleeding Esophageal Varices	R. B. Smith III, W. D. Warren	26:229-34

Low Renin Hypertension	J. C. Gunnells Jr., W. L. McGuffin Jr.	26:259-75
The Steal Syndromes	J. F. Toole, C. P. McGraw	26:321-29
The Total Correction of Congenital Heart Disease	E. Aberdeen	26:451-64
Pathophysiology of Heart Block	J. F. Moran, R. M. Gunnar	26:471-83
Medical and Surgical Treatment of the Wolff-Parkinson-White Sydrome	D. Durrer	27:63-68
Coronary Artery Spasm: A New Appraisal	A. A. Kattus	27:69-75
Medical Management of Myocardial Infarction	M. S. Klein, B. E. Sobel	27:89-108
Mitral Valve Prolapse	E. D. Wigle, H. Rakowski, N. Ranganathan, M. D. Silver	27:165-80
The Detection of Asymptomatic Coronary Artery Disease	V. F. Froelicher Jr.	28:1-12
Ionic Basis of Myocardial Contractility	G. A. Langer	28:13-20
Hypertrophic Cardiomyopathy	P. M. Shah	28:235-50
Critique of Coronary Artery Bypass Surgery	C. B. Mullins, K. Lipscomb	28:271-89
Development of Coronary Collateral Circulation	A. S. Wechsler	28:341-48
Prophylaxis of Infective Endocarditis: A Reevaluation	J. N. Sipes, R. L. Thompson, E. W. Hook	28:371-91
The Role of Ethanol in Cardiac Disease	T. J. Regan, P. O. Ettinger, B. Haider, S. S. Ahmed, H. A. Oldewurtel, M. M. Lyons	28:393-409
CLINICAL PHARMACOLOGY		
Drug Metabolism in the Newborn	M. Eriksson, S. J. Yaffe	24:29-40
Clinical Pharmacology of Drug Interactions	I. H. Raisfeld	24:385-418
Tardive Dyskinesia	G. W. Paulson	26:75-81
Clinical Pharmacology of Digitalis Glycosides	J. E. Doherty, J. J. Kane	26:159-71
Analgesic Abuse and Renal Disease	T. Murray, M. Goldberg	26:537-50
Rational Psychopharmacology for Patients with Medical Diseases	D. J. Greenblatt, R. I. Shader, S. Lofgren	27:407-20
Effective Use of Analgesic Drugs	L. E. Hollister	27:431-46
DERMATOLOGY		
Cytokinetics in Diseases of Epidermal Hyperplasia	G. D. Weinstein, J. L. McCullough	24:345-52
Skin Lipids and Acne	J. S. Strauss, P. E. Pochi, D. T. Downing	26:27-32
Human Disease Associated with Exposure to Light	J. H. Herndon Jr., R. G. Freeman	27:77-87
Current Problems in Mycosis Fungoides and Sezary Syndrome	R. K. Winkelmann, W. A. Caro	28:251-69
Pathophysiology of Psoriasis	J. J. Voorhees	28:467-73
ENDOCRINOLOGY AND METABOLISM		
Blood Levels of Parathyroid Hormone in Disorders of Calcium Metabolism	E. Reiss, J. M. Canterbury	24:217-32
Insulin Receptors	B. Desbuquois, P. Cuatrecasas	24:233-40
Human Prolactin	H. Friesen, P. Hwang	24:251-70
Recent Advances in Evaluation of Gonadotropic Hormones	J. L. Vaitukaitis, G. T. Ross	24:295-302
Normal and Abnormal Alpha Cell Function	R. H. Unger	24:303-8
Thyrotropin Releasing Hormone: Secretion and Actions	J. F. Wilber	24:353-64
Isolated Growth Hormone Deficiency and Related Disorders	T. J. Merimee	25:137-42
Serum Triiodothyronine in Man	R. D. Utiger	25:289-302
Vasopressin in Osmotic Regulation in Man	G. L. Robertson	25:315-22
Vitamin D Metabolism	M. F. Holick, H. F. DeLuca	25:349-67
Metabolic Response to Infection	W. R. Beisel	26:9-20
The Role of Hormone Receptors in the Action of Adrenal Steroids	D. Feldman	26:83-90
Regulation of Albumin Metabolism	M. A. Rothschild, M. Oratz, S. S. Schreiber	26:91-103
Muscle Amino Acid Metabolism and Gluconeo-		

genesis	N. B. Ruderman	26:245-58
Low Renin Hypertension	J. C. Gunnells Jr., W. L. McGuffin Jr.	26:259-75
Endocrine Therapy of Breast Cancer	W. L. McGuire	26:353-63
Interconversions of Apolipoprotein Fragments	C. B. Blum, R. I. Levy	26:365-71
Effects of Anesthesia on Intermediary Metabolism	E. A. Brunner, S. C. Cheng, M. L. Berman	26:391-401
Gonadotropin-Releasing Hormone	S. S. C. Yen	26:403-17
Relation Between Growth Hormone and Somatomedin	J. J. Van Wyk, L. E. Underwood	26:427-41
Active Form of the Thyroid Hormone and Somatomedin	S. H. Ingbar, L. E. Braverman	26:443-49
Hyperglycemia, Polyol Metabolism, and the Complications of Diabetes Mellitus	K. H. Gabbay	26:521-36
Potential Microwave Injuries in Clinical Medicine	D. I. McRee	27:109-15
The Effects of Calcium and Phosphorus in Hemodialysis	R. S. Goldsmith	27:181-90
Circadian Rhythms and Episodic Hormone Secretion in Man	E. D. Weitzman	27:225-43
Estrogens and the Human Male	R. Marcus, S. G. Korenman	27:357-70
Somatostatin: Physiological and Clinical Significance	R. Guillemin, J. E. Gerich	27:379-88
Medical Aspects of Riot Control (Harassing) Agents	J. P. Sanford	27:421-29
Impaired Peripheral Conversion of Thyroxine to Triiodthyronine	R. R. Cavalieri, B. Rapoport	28:57-65
Diagnosis of Hyperparathyroidism	H. G. Bone III, W. H. Snyder, C. Y. C. Pak	28:111-17
The Role of Glucagon in the Endogenous Hyperglycemia of Diabetes Mellitus	R. H. Unger, L. Orci	28:119-30
Paget's Disease of Bone	W. G. Ryan	28:143-52
Classification of the Hyperlipidemias	R. J. Havel	28:195-209

GASTROINTESTINAL TRACT

Action of Cholera Toxin on Fluid and Electrolyte Movement in the Small Intestine	G. W. G. Sharp	24:19-28
Enterotoxic Diarrheal Syndromes	S. B. Formal, H. L. Dupont, R. B. Hornick	24:103-10
Extraction of Insulin by Liver	J. B. Field	24:309-14
Role of Gastrin in Hypersecretory Disorders in Man	J. E. McGuigan	25:233-49
Regulation of Albumin Metabolism	M. A. Rothschild, M. Oratz, S. S. Schreiber	26:91-104
Pathogenesis of Hepatic Coma	L. Zieve, D. M. Nicoloff	26:143-57
Selective Distal Splenorenal Shunt for Bleeding Esophageal Varices	R. B. Smith III, W. D. Warren	26:229-34
Disorders of the Lower Esophageal Sphincter	R. S. Fisher, S. Cohen	26:373-90
Stress Ulceration in the Acutely Ill	J. J. Skillman, W. Silen	27:9-22
Infectious Particles in Hepatitis	A. P. Waterson	27:23-35
Parietal Cell Vagotomy Without Drainage for the Treatment of Duodenal Ulcer	P. H. Jordan Jr.	27:191-97
New Observations in Crohn's Disease	H. D. Janowitz, D. B. Sachar	27:269-85
Metabolic Complications of Jejunoileal Bypass Operations for Morbid Obesity	H. W. Scott Jr., R. H. Dean, H. J. Shull, F. W. Gluck	27:397-405
Viral Agents and Diarrheal Illness	H. L. DuPont, L. Portnoy, R. H. Conklin	28:167-77

HEMATOLOGY

Platelet Function Tests in Disease	T. B. Tschopp, M. B. Zucker	24:1-18
Radiation and Drug Therapies, and Leukemia	A. Potolsky, W. P. Creger	24:75-82
Current Concepts in the Treatment of Chronic Myelocytic Leukemia	J. E. Sokal	24:281-88
Nature and Biological Activities of Degradation Products of Fibrinogen and Fibrin	H. C. Kwaan, G. H. Barlow	24:335-44
2, 3-DPG and Erythrocyte Oxygen Affinity	G. J. Brewer	25:29-38

Treatment of Acute Granulocytic Leukemias	J. Bernard, M. Weil, C. Jacquillat	25:39-51
Mechanism of Action of Hemoglobin	A. Arnone	25:123-30
Immuno-Prevention of Rh Hemolytic Disease of the Newborn	R. B. McConnell, J. C. Woodrow	25:165-78
Marrow Transplantation for Aplastic Anemia and Acute Leukemia	P. Neiman, E. D. Thomas, C. D. Buckner, R. Storb, A. Fefer, H. Glucksberg, K. G. Lerner	25:179-98
Clinical Utility of In Vitro Evaluation of Granulopoiesis	P. L. Greenberg, S. L. Schrier	25:269-78
Molecular Immunology of Factor VIII	T. S. Zimmerman, T. S. Edgington	25:303-14
Control of Platelet Production	L. A. Harker	25:383-400
Intravascular Coagulation in Liver Disease	M. Verstraete, J. Vermylen, D. Collen	25:447-55
Von Willebrand's Disease	L. Holmberg, I. M. Nilsson	26:33-44
Diagnosis and Prevention of Postoperative Deep-Vein Thrombosis	L. P. Le Quesne	26:63-74
Cellular Kinetics, Staging, and Immunoglobulin Synthesis in Multiple Myeloma	B. G. M. Durie, S. E. Salmon	26:283-88
Granulocyte Transfusion Therapy	D. J. Higby, E. S. Henderson	26:289-306
The Antiglobulin Test in Autoimmune Hemolytic Anemia	W. F. Rosse	26:331-36
Thalassemia	B. G. Forget, D. G. Nathan	26:345-51
Hemochromatosis: Course and Treatment	S. Sherlock	27:143-49
Vitamin K and the Oral Anticoagulant Drugs	R. A. O'Reilly	27:245-61
Medical Management of Venous Thrombosis	S. Wessler	27:313-19
Factors in the Pathomechanism of Chronic Lymphocytic Leukemia	H. Theml, R. Love, H. Begemann	28:131-41
Paroxysmal Nocturnal Hemoglobinuria (PNH) as a Clonal Disorder	R. C. Hartmann, A. B. Arnold	28:187-94
Cold Agglutinins and Cold-Agglutinin Disease	M. M. Frank, J. P. Atkinson, J. Gadek	28:291-98
Treatment of Idiopathic Thrombocytopenic Purpura (ITP)	Y. S. Ahn, W. J. Harrington	28:299-309
INFECTION		
Preventive Antibiotic Management in Surgery	J. F. Burke	24:289-94
Venezuelan Equine Encephalitis Virus Infection in Man	N. J. Ehrenkrantz, A. K. Ventura	25:9-14
Food Poisoning Due to Vibrio Parahaemolyticus	W. H. Barker Jr., E. J. Gangarosa	25:75-81
Preventive Therapy of Tuberculosis	D. L. Brummer	25:115-22
Infectious Complications of Nonalcoholic Drug Abuse	D. B. Louria	25:219-31
Antiviral Chemotherapy	J. P. Luby, M. T. Johnson, S. R. Jones	25:251-67
Metabolic Response to Infection	W. R. Beisel	26:9-20
Granulocyte Transfusion Therapy	D. J. Higby, E. S. Henderson	26:289-306
Urinary Tract Symptoms and Infections	J. P. Sanford	26:485-98
Biology of Endotoxin	R. J. Elin, S. M. Wolff	27:127-40
Significance of Serum Protein and Tissue Binding of Antimicrobial Agents	W. A. Craig, C. M. Kunin	27:287-300
Immunoprophylaxis of Gram-Negative Bacillary Infections	W. R. McCabe	27:335-41
Reye's Syndrome	G. M. Schiff	27:447-52
Cytomegalovirus (CMV) in the Compromised Host(s)	R. F. Betts, J. B. Hanshaw	28:103-10
Pertussis in the Adult	C. C. Linnemann Jr., J. Nasenbeny	28:179-85
The Emergence of Group B Streptococci in Infections of the Newborn Infant	B. F. Anthony, D. M. Okada	28:355-69

KIDNEYS AND URINARY TRACT		
Urethrovesical Dysfunction and Infection	F. Hinman	24:83-88
Renal Energy Metabolism and Sodium Reabsorption	P. R. Lewy, A. Quintanilla, N. W. Levin, R. H. Kessler	24:365-84
Immunopathology and Glomerulonephritis	C. B. Wilson, F. J. Dixon	25:83-98
Calcium Oxalate Renal Stones	E. L. Prien Jr.	26:173-79
Low Renin Hypertension	J. C. Gunnells Jr., W. L. McGuffin Jr.	26:259-75
Acute Renal Failure (Vasomotor Nephropathy): Micropunture Studies of the Pathogenetic Mechanism	D. E. Oken	26:307-19
Urinary Tract Symptoms and Infections	J. P. Sanford	26:485-98
Analgesic Abuse and Renal Disease	T. Murray, M. Goldberg	26:537-50
Late Sequelae of Poststreptococcal Glomerulonephritis	D. S. Baldwin, R. G. Schacht	27:49-55
Glomerular Injury in Patients with Neoplasia	B. S. Kaplan, J. Klassen, M. H. Gault	27:117-25
Interrelations of Renin, Angiotensin II, and Sodium in Hypertension and Renal Failure	H. Gavras, J. A. Oliver, P. J. Cannon	27:485-521
Glomerular Filtration of Fluid and Macromolecules: The Renal Response to Injury	D. A. Maddox, B. M. Brenner	28:91-102
Renal Function During and Following Obstruction	D. R. Wilson	28:329-39
Urate Excretion	G. M. Fanelli Jr.	28:349-54
MISCELLANEOUS		
Chromosome Identification	K. Hirschhorn	24:67-74
Genetic Aspects of Gout	M. A. Becker,, J. E. Seegmiller	25:15-28
Medical Aspects of Recreational Skin and Scuba Diving	J. P. Sanford	25:401-10
Dental Caries	R. J. Gibbons, J. van Houte	26:121-36
Unusual Etiologies of Fever and Their Evaluation	S. M. Wolff, A. S. Fauci, D. C. Dale	26:277-81
Implications of Mass Automated Instruments on Medical Practice	R. R. Grams	27:199-206
Confidentiality in Medical Practice	M. Grossman	28:43-55
MUSCULO-SKELETAL DISORDERS		
Crystal Deposition Joint Disease	D. J. McCarty	25:279-88
NEOPLASIA		
Scanning Techniques to Assess Thryoid Nodules	J. L. Quinn III, R. E. Henkin	26:193-201
Treatment of Testicular Cancer	D. I. Twito, B. J. Kennedy	26:235-43
Cellular Kinetics, Staging, and Immunoglobulin Synthesis in Multiple Myeloma	B. G. M. Durie, S. E. Salmon	26:283-88
Endocrine Therapy of Breast Cancer	W. L. McGuire	26:353-63
The Role of Radiation Therapy in the Definitive Treatment of Adenocarcinoma of the Prostate	G. R. Ray, M. A. Bagshaw	26:567-88
Tumor Antigens and Tumor-Host Relationships	R. W. Baldwin, M. R. Price	27:151-63
The Functions of the Macrophage in Malignant Disease	P. Alexander	27:207-24
Multiple Endocrine Neoplasia Type II	S. A. Wells Jr., D. A. Ontjes	27:263-68
The Role of Oncogenic Viruses in Neoplasia	P. A. Klein, R. T. Smith	28:311-27
Surgical Adjuvant Chemotherapy	M. Perloff, J. F. Holland	28:475-88
NEUROLOGY		
The Lesch-Nyhan Syndrome	W. L. Nyhan	24:41-60
Metabolism of Ketone Bodies by the Brain	L. Sokoloff	24:271-80
The Minimal Brain Dysfunction Syndrome	P. H. Wender	26:45-62
Tardive Dyskinesia	G. W. Paulson	26:75-81
Pathogenesis of Hepatic Coma	L. Zieve, D. M. Nicoloff	26:143-57
The Steal Syndromes	J. F. Toole, C. P. McGraw	26:321-29
Huntington's Disease	I. Shoulson, T. N. Chase	26:419-26
A Definition of Cerebral Death	B. Boshes	26:465-70
Neuroregulators and Sleep Mechanisms	R. B. Holman, G. R. Elliott, J. D. Barchas	26:499-520
Subacute Sclerosing Panencephalitis	R. S. Johannes, J. L. Sever	26:589-601
Temporal Lobe Epilepsy: Neurological and		

Behavioral Aspects	I. Sherwin	27:37-47
NUTRITION		
Modification of Brain Serotonin by the Diet	J. D. Fernstrom	25:1-8
Applications and Hazards of Total Intravenous Hyperalimentation	S. J. Dudrick, J. M. Long III	28:517-28
ONCOLOGY		
Current Concepts in Treatment of Lymphomas with Special Reference to Hodgkin's Disease	L. Stutzman	24:325-34
Medical Management of the Hypercalcemia of Malignancy	L. J. Deftos, R. Neer	25:323-31
The Cell Surface and Malignant Transformation	R. E. Pollack, P. V. C. Hough	25:431-46
OPHTALMOLOGY AND OTORHINOLARYNGOLOGY		
Corneal Transplantation	S. L. Forstot, H. E. Kaufman	28:21-35
PSYCHIATRY		
Behavioral Aspects of Chromosomal Disorders	S. Kessler, R. H. Moos	24:89-102
Mental Effects of L-Dopa	D. L. Murphy	24:209-16
Pharmacotherapy of Schizophrenia	G. L. Klerman	25:199-217
Biogenic Amines and Affective Disorders	J. J. Schildkraut	25:333-48
X-Linkage in Manic-Depressive Illness	R. J. Cadoret, G. Winokur	26:21-25
Biochemical Aspects of Schizophrenia	S. Matthysse, J. Lipinski	26:551-65
Reassement of Homosexuality and Transsexualism	R. C. Friedman, R. Green, R. L. Spitzer	27:57-62
The Relation of Psychological Stress to Onset of Medical Illness	M. W. Hurst, C. D. Jenkins, R. M. Rose	27:301-12
Behavioral and Biochemical Interrelations in Alcoholism	J. H. Mendelson, N. K. Mello	27:321-33
Psychological Assistance for the Dying Patient and His Family	T. P. Hackett	27:371-78
Neuroendocrine Dysfunction in Depressive Illness	E. J. Sachar	27:389-96
The Sleep Apena Syndromes	C. Guilleminault, A. Tilkian, W. C. Dement	27:465-84
New Concepts in the Diagnosis of Psychiatric Disorders	E. Robins	28:67-73
PUBLIC HEALTH AND ENVIRONMENTAL MEDICINE		
Methadone Treatment of Heroin Addiction	D. X. Freedman, E. C. Senay	24:153-64
RADIOLOGY		
Diagnosis and Prevention of Postoperative Deep-Vein Thrombosis	L. P. Le Quesne	26:63-74
Scanning Techniques to Assess Thyroid Nodules	J. L. Quinn III, R. E. Henkin	26:193-201
The Role of Radiation Therapy in the Definitive Treatment of Adenocarcinoma of the Prostate	G. R. Ray, M. A. Bagshaw	26:567-88
REPRODUCTION		
Estimation of Gestational Age in the Fetal State	J. W. C. Johnson Jr., G. B. Odell	24:165-74
Gonadotropin-Releasing Hormone	S. S. C. Yen	26:403-17
Some Implications of Population Control	A. C. Barnes	27:1-7
RESPIRATORY SYSTEM		
Mechanics of Air Flow in Airway Obstruction	A. C. DeGraff Jr., A. Bouhuys	24:111-34
Endogenous Carbon Monoxide Metabolism	R. F. Coburn	24:241-50
Oxygen Toxicity	W. G. Wolfe, W. C. DeVries	26:203-17
The Adult Respiratory Distress Syndrome	P. C. Hopewell, J. F. Murray	27:343-56
Hypersensitivity Pneumonitis: Current Concepts of Etiology and Pathogenesis	M. Lopez, J. Salvaggio	27:453-63
Diagnosis and Treatment of Fat Embolism	J. A. Moylan, M. A. Evenson	28:85-90
Diagnosis of Pulmonary Embolism	G. V. R. K. Sharma, D. E. Tow, A. F. Parisi, A. A. Sasahara	28:159-66
Pulmonary Complications of General Surgery	A. K. Pierce, J. Robertson	28:211-21
THERAPEUTICS AND TOXICOLOGY		
Drug Metabolism in the Newborn	M. Eriksson, S. J. Yaffe	24:29-40
Clinical Pharmacology of Drug Interactions	I. H. Raisfeld	24:385-418
The Effects of Fluorinated Anesthetics on the Liver and Kidneys	L. S. Gottlieb, C. Trey	25:411-29